

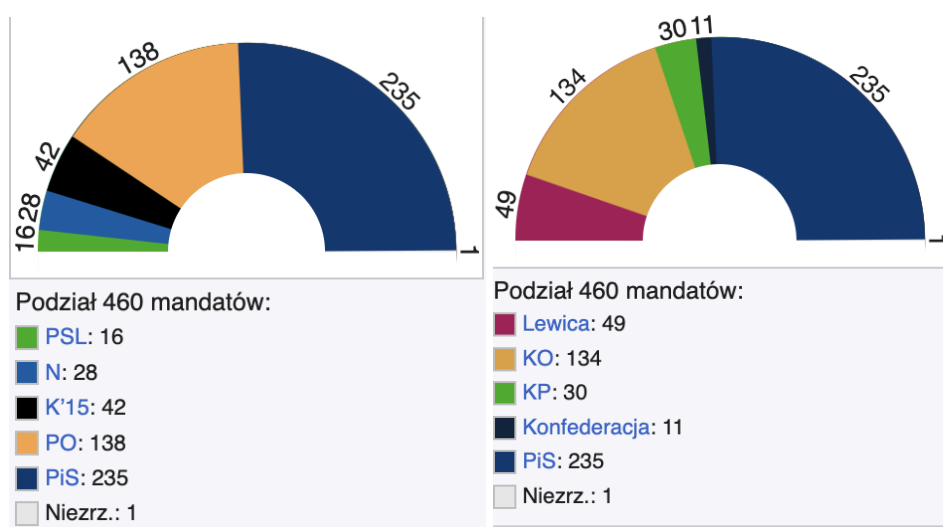
CLARIN - OPIS DANYCH i INSTRUKCJA

Zbiór danych:

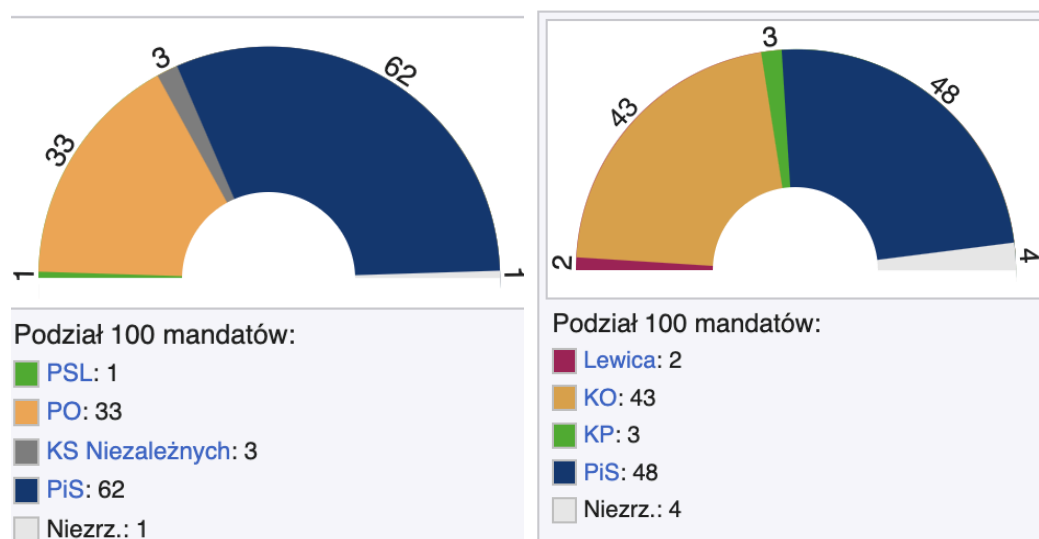
Zbiór danych obejmuje wypowiedzi pochodzące z rodziny korpusów ParlaMint. Wykorzystaliśmy wersję 3.0 stanowiącą wielojęzyczny zestaw 26 korpusów zawierających debaty parlamentarne rozpoczynające się głównie w 2015 r. i trwające do połowy 2022 r. Każdy z korpusów zawiera od 9 do 125 milionów słów i informacje dotyczące prelegentów (imię i nazwisko, płeć, status posła, przynależność partyjna, koalicja partyjna/opozycja). Korpusy są podzielone na terminy, sesje i spotkania. Transkrypcje wypowiedzi posłów i senatorów dodatkowo zawierają odpowiednio oznaczone komentarze padające z publiczności oraz informacje o przerwach w transkrypcji np. brawach itp.

Do anotacji postanowiliśmy wykorzystać wypowiedzi z lat 2016 i 2022 żeby zapewnić różnorodność i uwzględnić wypowiedzi posłów i senatorów z różnych kadencji. Analizowane przez nas wypowiedzi pochodzą zatem z posiedzeń posłów z VIII i IX kadencji (Marszałkowie Sejmu RP: Marek Kuchciński/Elżbieta Witek i Elżbieta Witek) oraz senatorów z IX i X kadencji (Marszałkowie Senatu RP: Stanisław Karczewski i Tomasz Grodzki). Podział mandatów widoczny jest na Rysunku 1.

Podział mandatów w 2016 roku:



Podział mandatów w 2022 roku:



Rysunek 1. Podział mandatów w Sejmie i Senacie w latach 2016 i 2022..

Ze względu na duże rozmiary korpusów postanowiliśmy dokonać preanotacji, czyli wstępnego przetworzenia tekstów narzędziami automatycznymi, w celu wykrycia części interesujących nas zjawisk, które następnie zostały poddane anotacjom ręcznym. W tym celu zastosowaliśmy dwa podejścia - w pierwszym wypowiedzi mogące zawierać przejawy agresji zostały wyodrębnione poprzez oznaczenie negatywnego sentymentu natomiast w drugim zastosowaliśmy listę słów.

1. Analiza sentymentu

We wszystkich wypowiedziach posłów i senatorów z 2016 roku automatycznie oznaczono pozytywny i negatywny sentyment. W tym celu wykorzystaliśmy słownik zawierający 5421 słowa pozytywne i negatywne uzyskane z dwóch źródeł. Pierwszym źródłem był plWordNet Emo, który zawiera 32 tysiące jednostek leksykalnych opatrzonych ręcznie etykietami na poziomie znaczeń słów (Zaśko-Zielińska i in., 2015) Drugie źródło to ręcznie oznaczony słownik zawierający 1774 słów negatywnych i 1493 pozytywnych, pobrany z analizatora sentymentu opartego na regułach (Buczyński i Wawer, 2008). Słownik został stworzony ręcznie i odzwierciedla sentyment najczęstszego znaczenia słowa, dlatego nie wymaga ujednolicania znaczenia słowa.

Aby uwzględnić długość wypowiedzi, zastosowaliśmy skalowaną liczbę negatywnych słów czyli wskaźnik będący stosunkiem tokenów negatywnych do wszystkich tokenów. Według tej miary uszeregowaliśmy wszystkie wypowiedzi z 2016 roku i wybraliśmy do anotacji 210 najlepszych wypowiedzi.

Kody wykorzystywane do preselekcji wypowiedzi znajdują się w pliku `perpare_corpus_for_annotation.ipynb`.

Ze względu na dużą liczbę wypowiedzi, które mimo negatywnego sentymentu nie zawierały przykładów agresji postanowiliśmy zastosować podejście oparte o listę słów.

2. Lista słów

W oparciu o analizę wypowiedzi posłów i senatorów z lat 2016-2017 oraz artykułów dotyczących agresji językowej w komunikacji dydaktycznej (Krzyżyk, 2023) oraz w polityce (Fastyn, 2013; Grunwald 2020) opracowaliśmy listę najczęściej pojawiających się słów związanych z agresją w podziale na kategorie oraz pozostałymi oznaczanymi przez nas cechami wypowiedzi wpływającymi na ich wiarygodność.

Słowa związane z agresją::

- odnoszące się do (braku) kompetencji: np. niekompetencja, potrafić, umieć, nieudolny
- związane z insynuowaniem kłamania: np. kłamstwo, kłamać, bzdura, krętactwo, prawda
- odnoszące się do zachowań: nadużywać, marnować, sprzedać, wstydzić
- odnoszące się do cech: arogancja, ignorancja
- związane z deprecjonowaniem: kpina, żart, skandal, burdel, chaos
- wyzwiska: wariat, komunista, koleś
- uogólnienie: wszyscy, zawsze, nigdy
- antagonizm, podział na "my" i "oni": wy, wasz/wasze/wasza
- związane z tematami wzbudzającymi emocje: LGBT, aborcja

Słowa związane z pewnością/przekonaniem o słuszności: stanowczo, absolutnie, pewnośc, oczywiście, jakby

Słowa związane z powoływaniem się na autorytet: wyniki, statystyki, badania, według, zgodnie,

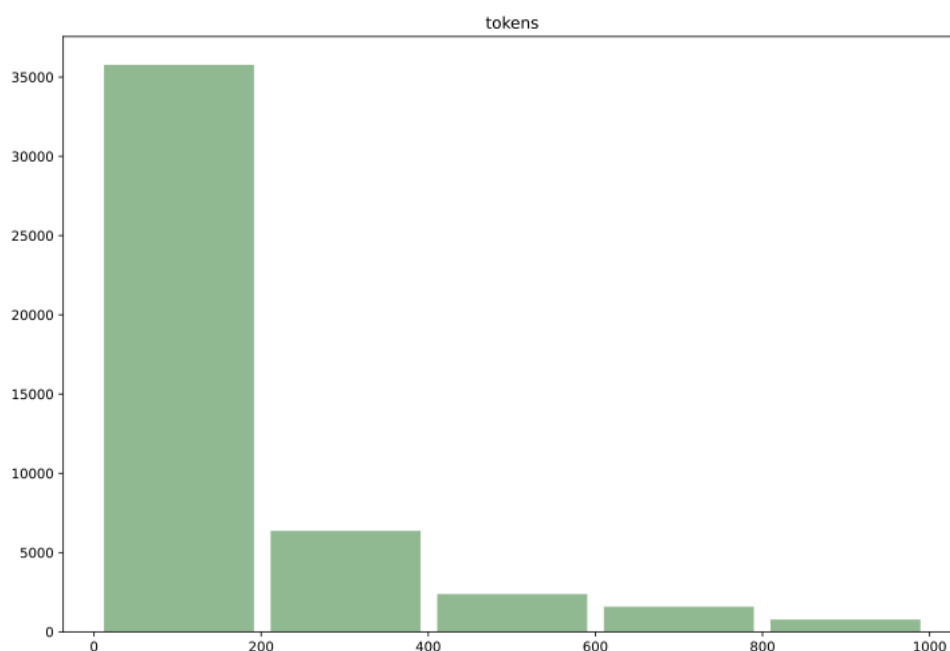
Słowa świadczące o relatywizmie: czasami, strony, jednocześnie, wypadek, wyjątek

Pełna lista słów dostępna jest w osobnym pliku (clarin_lista_slow.xls) a bardziej szczegółowy opis kategorii wraz z przykładami w instrukcji do anotacji.

Strategie preselekcji - porównanie

Zdecydowaliśmy się usunąć ze zbioru danych wypowiedzi najkrótsze (długość mniejsza niż 200 tokenów), które stanowiły zwykle podziękowania lub udzielenie głosu przez Marszałka oraz wypowiedzi najdłuższe (długość powyżej 1000 tokenów), które z kolei były zazwyczaj przedstawieniem stanowiska klubu w sprawie ustawy/projektu.

Liczebności wypowiedzi określonych długości z roku 2016 (po zastosowaniu filtrów) widoczne są na histogramie (Rysunek 2.).



Rysunek 2. Histogram prezentujący długości wypowiedzi.

W losowo wybranych wypowiedziach z 2016 roku, nie filtrowanych w żaden sposób, przejawy agresji zostały oznaczone w 31 wypowiedziach z 150 co stanowi 21%. Po przefiltrowaniu wypowiedzi pod kątem występowania słów ze słownika negatywnego sentymentu (NEG) w 23% z nich stwierdziliśmy przejawy agresji. Wreszcie, korzystając z dedykowanej listy słów, znaleźliśmy interesujące nas cechy (przejawy agresji, przykłady odwoływania się do autorytetów, wyrażenia wskazujące na pewność mówiącego lub jej brak oraz przykłady względności wypowiedzi) aż w 39% wypowiedzi. Wyniki podsumowuje Tabela 1.

Strategia preselekcji	Stosunek
brak filtrów	21%
NEG sentyment	23%
lista słów	39%

Tabela 1. Skuteczność strategii preselekcji.

Anotacjom poddaliśmy 638 wypowiedzi - 48 pochodziło ze zbioru wyłonięgo za pomocą negatywnego sentymentu spośród wypowiedzi posłów i senatorów z 2016 roku, 90 z tego samego roku, ale wyselekcjonowanych za pomocą słownika a pozostałe 500 to wypowiedzi wyselekcjonowane przy użyciu metody słownikowej spośród wypowiedzi pochodzących z 2022r.

Instrukcja

Zadaniem anotatorów jest oznaczenie w sprawozdaniach stenograficznych posiedzeń Sejmu i Senatu przejawów agresji werbalnej, czyli wszystkich fragmentów wypowiedzi wypowiadanych z intencją urażenia, obrażenia lub poniżenia adresata, podważenia jego kompetencji, poddania krytyce jego światopoglądu, dokonań i zachowań.

Za przejawy agresji werbalnej należy uznać: napastliwy ton wypowiedzi, zastosowanie środków leksykalnych w postaci obelg, wulgaryzmów, wyzwisk, krytyki, wyrażen ocenających o wyraźnym zabarwieniu negatywnym. Elementy językowe składające się na agresję werbalną są nierozzerwalnie związane z sytuacją i kontekstem wypowiedzi. Pod względem aksjologicznym- zawsze z doświadczeniem negatywnym.

Dodatkowo oznaczeniom podlegają charakterystyki wypowiedzi, które mogą wpływać na jej wiarygodność - powoływanie się na wyniki badań i opinie autorytetów, oznaki przekonania posłów w słuszność własnych wypowiedzi oraz względność czyli dopuszczanie pewnej niejednoznaczności i dostrzeganie zarówno mocnych, jak i słabych stron proponowanych rozwiązań.

Poniżej znajdują się przykłady skategoryzowanych komunikatów o charakterze agresywnym, wyrażonych w sposób bezpośredni lub pośredni.

1. Inwektywy opisowo-oceniające.

Obrażliwe, poniżające określenia, słownictwo znieważające o wydźwięku jednoznacznie pejoratywnym. Najczęściej występujące podtypy:

- odzwierzęce
np. baran, osioł, świnia, gad, leszcz, szczwany lis
- odnoszące się do stanu umysłu (sugerujące ułomność psychiczną)
np. nienormalni, wariat, idiota, debil
- analogie historyczne
np. łże-elita, komuniści, postkomuniści, kolaboranci, faszyści, nacjonaści, ubecy
- kalumnie – krzywdzące oskarżenia szkalujące kogoś, niszczące jego dobre imię
np. kłamcy, patałachy, tchórze, nieroby, niepraw i niesprawiedliwi, partacze,

2. Podważanie prawdomówności i manipulowanie informacją.

- kłamanie

np. przestańcie kłamać, kłamstwo, oszukaliście Polaków, nieprawda, okłamaliście nas, herezje, zwykła fikcja, bujda, humbug

- celowe zatajenie prawdy

np. przemilczeliście najważniejsze, nie powiedzieliście całej prawdy, apeluję o prawdę, mistyfikacja

- wprowadzanie w błąd

np. wprowadzacie w błąd, mijacie się z prawdą, opowiadacie bajki, przeinaczacie fakty,

3. Antytetyczny podział na “my - oni”.

Podkreślanie różnic pomiędzy koalicją rządzącą a opozycją, zwykle za pomocą zaimka osobowego „my” i dzierżawczego „naszych”, często wzmocnione poprzez zastosowanie leksemów ‘zawsze’ i ‘nigdy’. W ten sposób przedstawiane są osiągnięcia własnej partii i dyskredytowane działania, dokonania i przekonania partii oponentów.

- podkreślanie różnic

np. my w przeciwieństwie do was jesteśmy zawsze otwarci na dialog, to wy zachęcaliście Polaków do brania kredytów we frankach a my musimy teraz po was sprzątać,

- dyskredytowanie oponentów

np. wasze zaniechania, wasza nieudolność, wasza wygodność i przyjemności, Dla was rządzenie polega właśnie na tym, na załatwianiu interesów własnych i własnej partii, wy nie mogliście o to zadbać, Wasza Polska to Polska chaosu, obłudy i niekompetencji, inflacji i drożyzny.

4. Pozorna grzeczność (inaczej antyetykieta).

Krytyczne wypowiedzi zachowujące pozory grzeczności, których właściwą illokucją jest obraza - np. zarzucenie braku wiedzy. Obejmują one nieprzychylną opinię, przytyki, uszczypliwości, często ukryte pod pozornie przychylnym komentarzem:

- zwroty grzecznościowe o zabarwieniu ironicznym

np. Gratuluję tej głębokiej analizy, Szanowni państwo, czytajcie ze zrozumieniem, Czcigodna pani posłanka, bardzo się cieszę, że zaczęliście Państwo czytać projekty, Wybaczcie mi .., ale..., Ja bardzo przepraszam...

- fałszywe komplementy. Pogarda lub kpina wyrażona za pomocą pozornej aprobaty

np. Genialne wystąpienie, genialne., Pani Marszałek! Wysoki Sejmie! Pani Minister! Dokonała pani dzisiaj rzeczy niebywałej. Z charakterystyczną dla siebie arogancją i butą obraziła pani uczniów, nauczycieli, rodziców, a nawet samorządy.

- zawołowana krytyka i fałszywa troska

np. ja sobie nie przypominam, żebyście kiedyś..., ja nie chciałbym naruszać programu, ale...

5. Atak na kompetencje.

Zarzuty dotyczące braku wiedzy i umiejętności kluczowych dla rządzenia i realizowania obietnic wyborczych. Występują w nich najczęściej bardzo ekspresyjne sformułowania o charakterze napastliwym, demaskatorskim czy wręcz obraźliwym. Ich wyraźne i zamierzone nasycenie sądami wartościującymi negatywnie ma na celu zdyskredytowanie kompetencji przeciwników politycznych i podważenie ich wiarygodności w oczach obywateli.

- krytyka dotychczasowych dokonań

np. *złe rozwiązanie, Ustawa – absurdalna, źle przygotowana, bubel prawny, szkodliwa, nieprzydatna, pisana na kolanie; bubel prawny, bubel bubleń pogania, jedna wielka katastrofa, gigantyczny chaos, polski nieład, Nowy Chaos, projekt nadaje się do kosza, Zrealizowane obietnice to karykatury waszych planów,*

- deprecjonowanie dokonań i dorobku wyrażone wprost w w. 2. osobie liczby mnogiej; wzmocnione poprzez zastosowanie partykuły “nie” lub w formie bezosobowej

np. *nie poradziliście sobie, nie potraficie, nie zrobiliście nic, nie umiecie nic w tym zakresie, kompletnie; nie potraficie gospodarować, bredzicie; Łatwo trwoni się cudze pieniądze;*

- wyrażenie dezaprobaty i nagany poprzez zastosowanie dla opisu semantycznego przymiotników wartościujących negatywnie

np. *nieudolny, partacki rząd; skrajnie niekompetentny, niegospodarny, opieszale rząd, totalna beczynność rządu, gigantyczny chaos, absolutnie skandaliczne*

- ...

6. Antydialogiczne uogólnienia

Strategia uogólnień z ich leksemowymi wykładnikami, takimi jak *każdy, wszyscy, wszędzie, zawsze, nigdy, zwykle* itd.

- krytyka oponentów

np. *nikt za nic nie odpowiadał, nigdy nie słuchacie sugestii*

- chwalenie własnej partii

np. *my zawsze przychodzimy na spotkania, dla nas wszystkie głosy są ważne*

- solidaryzowanie się z przeciętnym Polakiem/obywatelem

np. *dzięki naszym zmianom wszyscy wreszcie poczujemy się bezpiecznie, Za ich decyzje zapłacimy my wszyscy.*

7. ironia

Ukryta pod pozorem uprzejmości i wyrażona niekiedy w sposób pozornie zabawny forma negatywnej oceny sytuacji, osoby czy jej kompetencji. Ironiczna pochwała jest na poziomie komunikacyjnym wyrazem krytyki, niezadowolenia, zawiedzionych oczekiwań i

rozczarowania. W komunikatach językowych tego typu dominantę stanowi pozorność aprobaty ze strony nadawcy w stosunku do odbiorcy i/lub jego zachowań.

- ocena osoby i kompetencji

np. Pani poseł polonistko, znawczyni finansów publicznych...

- ocena sytuacji

np. No wspaniale, po prostu kraina miodem i mlekiem płynąca

- ironiczna pochwała

np. Genialne wystąpienie, genialne; To doprawdy wspaniały pomysł!; głupota z dyktando

8. Zalecenia, rekomendacje, rady, nakazy, pouczenia.

Przybiera czasem formę pozornej grzeczności i troski, w rzeczywistości przekaz taki jest formą zawołanej krytyki i dominacji; zawiera sugestię określonej ułomności lub braku kompetencji. Ma charakter oceniająco-emocjonalny. Nieżyczliwe pouczanie ma za zadanie wywołać u odbiorców postawę posłuszeństwa. Konstrukcje językowe tego typu wyrażone są najczęściej w trybie rozkazującym. Uwaga: Użycie formy rozkaznika może być podkreślone graficznie wykrzyknikiem.

Proszę sobie uprzejmie jeszcze raz przesłuchać.

Szanowni państwo, czytajcie ze zrozumieniem

Proszę mi nie przerywać!

Proszę nie przeszkadzać.

Ale niech pan zejdzie z tej mównicy!

Jeżeli rząd zamierza podjąć czynności w tym zakresie, to proszę wskazać, jakie to czynności są planowane, co i w jakim terminie będziecie robić.

Chcę powiedzieć, żebyście państwo wzięli się za robotę, a nie za propagandę.

Z tej ścieżki musicie zejść

Zajmijcie się tym!

9. Pytania retoryczne

Nie stanowią elementu polemiki politycznej. Zadawane w napastliwym tonie z mównicy mają wywołać określoną reakcję adresata; zdeorientować go, zdyskredytować i zwrócić uwagę słuchaczy, podkreślić postawę, stanowisko autora komunikatu i nakierować odbiorców na konkretny tor myślenia, a przede wszystkim wzmocnić siłę wypowiedzi. Stanowią formę perswazji i umożliwiają manipulację odbiorcą (ułatwiając przyznanie racji czy pozyskanie aprobaty). i ewentualne podtypy

Co pani na to, pani Minister?

Co pan na to, panie Ministrze?

W jakim świecie pan poseł żyje?
Ale czy pani poseł wie, co pani mówi?
I mam pytanie: Co z tym zrobicie? Co z tym zrobicie?
Co tu się wydarzyło, kochani rządzący?
Na co wy jeszcze czekacie, do jasnej cholery?
Co państwo tak naprawdę proponujecie?
Skąd macie kasę?
Co z tym zrobiliście?
A co pan udźwignął?
Czy wy sobie z tym poradzicie?
Czy odbyła się pogłębiona dyskusja jak się do tego odnieść?
Dlaczego państwo udajecie fikcję niezależności?

10. Uszczypliwe żarty, kpina, drwina, szyderstwo.

Nacechowane w ten sposób komunikaty językowe są działaniami bezpośrednimi o wyrażnie pogardliwym wydźwięku, dyskredytującymi oponenta bez semantycznego zawoalowania,. Ich adresat zostaje wyrażnie ośmieszony bez łamania konwencji reguł grzecznościowych. Wykorzystanie elementów żartu zmienia rejestr komunikacyjny z poważnego na żartobliwy w formie, przy czym w dalszym ciągu stanowiący niepozbawioną złośliwości krytykę oponentów.

definicja, podtypy i przykłady

Miesiąc temu, jak rozmawialiśmy o tej ustawie, żartowaliśmy z kolegami, że pewnie zaraz dopiszą telewizor, a może jeszcze playstation.

Słyszałem do tej pory o diecie cud, ale ustawy cud to nie widziałem.

Niestety opozycja inspirowa i sufluje tę papkę intelektualną i konkuruje z braćmi Grimm i Krasickim w opowiadaniu bajek.

I naprawdę Bóg jeden wie, do czego nas to doprowadzi i co z tym okresem pomiędzy, a może nawet Bóg tego nie wie, bo filozofowie twierdzili, że nawet dla Boga jest niemożliwe uczynić niebyłym to, co już się wydarzyło. A w tym Sejmie to się dzieje.

A pani dostaje jeszcze jakieś role w teatrze czy tylko tutaj?

11. Głosy z sali:

Nacechowane emocjonalnie, zazwyczaj krótkie komentarze z sali sejmowej, wtrącane zwykle w trakcie czyjejś wypowiedzi. Mogą uniemożliwiać dokończenie wypowiedzi.

Uwaga: wykładnikiem illokucyjnym kpiny, szyderstwa i ironii w wypadku głosów z sali jest również śmiech. Sposobem jego wyrażania jest bezpośrednia werbalizacja w postaci zapisu "Ha ha ha", "Hahaha" lub "HAHAHA".

Zejdź z mównicy, Skandal, To nieprawda, Skończ już, Antysemita chwali, To już nudne jest, Naucz się czytać, Użyj mózgu, odpustowy kaznodzieja, mikrofon wszystko wytrzyma, skończ już z tymi bzdurami, Po co peplasz?

Powoływanie się na autorytet:

Uzasadnianie swojego stanowiska poprzez powołanie się na wyniki badań lub opinie osób uznawanych za autorytet.

- autorytet

W podobnym tonie wypowiadał się prof. Zbigniew Lew-Starowicz, autorytet jeśli chodzi o badanie tych wszystkich niekorzystnych zjawisk, także związanych z przestępczością na tle seksualnym

- badania

Bo, szanowni państwo, według badań szwajcarskich wartość funkcji lasu, wartość produkcyjna - pani dr Kissling-Näf ze Szwajcarii tak to określa - wynosi 5% ogólnej wartości ekosystemu, jaką daje nam las.

Badanie CBOS-u pokazywało, że 46% Polaków pozytywnie ocenia wymiar sprawiedliwości. Proponujecie nam zmianę, a już przekonaliśmy się na podstawie np. badań PISA z 2000 r., że nie sprawdził się tamten system.

Pewność

określenia mające świadczyć o stopniu przekonania posłów do wypowiedzianych przez siebie opinii.

Zresztą jestem absolutnie przekonany, że wszędzie w Polsce uzyskalibyśmy pozytywną odpowiedź na tak zadane pytanie: Czy jesteś za utworzeniem nowego uniwersytetu?

Dzisiaj mediami publicznymi, telewizją publiczną zarządza aparatczyk polityczny Jacek Kurski, i to jest bezsprzeczne.

Zgadzam się z tym absolutnie. Tu nie ma wątpliwości.

Względność

Dostrzeganie niejednoznaczności i zależności od różnych czynników. Zwykle przejawia się w dostrzeganiu zarówno zalet jak i wad danego rozwiązania czy projektu. W krytycznych opiniach mogą pojawiać się także dobre strony danego rozwiązania/projektu. Kategoria ta obejmuje także dopuszczanie pewnych wyjątków od proponowanych rozwiązań.

- wyjątek

Aborcja jest zabójstwem nienarodzonego dziecka i jako taka powinna być zabroniona, z wyjątkiem poważnych zagrożeń dla zdrowia i życia matki

- pozorna aprobata

To bardzo dobra ustawa, panie pośle Sawicki, i my poprzemy ją jako klub parlamentarny Kukiz'15, jednak nie rozwiązuje kompleksowo problemu sprzedaży bezpośredniej

Trzeba zauważyć że wasze intencje są dobre, jednak wykonanie pozostawia dużo do życzenia

- niejednoznaczność

Należy zauważyć, że tendencja od kilku lat jest wyraźnie spadkowa, jednak ta statystyka nie oddaje rzeczywistości w pełni.

Aspekty techniczne

Anotacje dokonywane są na poziomie zdań. W przypadku zdań wielokrotnie złożonych możliwe jest oznaczenie wyłącznie frazy zdania podrzędnego. Wyjątek stanowią głosy z sali, które są zwykle pojedynczymi wyrazami lub równoważnikami zdań.

Anotacji podlegają wszystkie znaki interpunkcyjne w obrębie zdania lub frazy zdania podrzędnego. Szczególną uwagę należy zwrócić na użycie formy rozkaznika podkreślone graficznie wykrzyknikiem.

Anotacji nie podlegają komunikaty dwu- lub wieloznaczne.