

MATEUSZ FLONT, WOJCIECH KAJTOCH, MARCIN PIELUŻEK
Uniwersytet Jagielloński

Perspektywy leksykalnej ilościowej analizy zawartości

Poniższy artykuł jest próbą zastosowania do badań języka dyskursu — klasycznej (zob.: Berelson 1952) metody prasoznawczej, zwanej analizą zawartości, definiowanej jako zespół

różnych technik systematycznego badania strumieni lub zbiorów przekazów, polegającego na możliwie obiektywnym (w praktyce zwykle: intersubiektywnie zgodnym) wyróżnianiu i identyfikowaniu ich możliwie jednoznacznie skonkretyzowanych, formalnych lub treściowych, elementów oraz na możliwie precyzyjnym (w praktyce zwykle: ilościowym) szacowaniu rozkładu występowania tych elementów i na głównie porównawczym wnioskowaniu, a zmierzającego przez poznanie zawartości przekazów do poznania innych elementów i uwarunkowań procesu komunikacyjnego (Pisarek 1983, 45).

Jej procedury pozwalały odpowiadać na pytanie o ranking silnie sfragmentowanych treści, proponowanych przez obfite strumienie tekstów i obrazów, którymi nieustannie „bombardują” nas media. Ponieważ cały strumień, nawet w wypadku jednej gazety, okazywał się niemożliwy do przebadania, należało pobrać z niego próby na tyle reprezentatywne, aby na ich podstawie można było sądzić o całości, a następnie wyodrębnić w tych próbach mierzalne elementy (tzw. jednostki analizy — teksty, akapity, sceny, postaci, motywy tematyczne itd.), zdecydować w jakich jednostkach będziemy je mierzyć (jakie będą tzw. jednostki pomiaru — np. ilość wystąpień, centymetry kwadratowe łamów im poświęcone, albo sekundy czasu antenowego) i jak wyznaczać ich ważność w całym przekazie.

Już jednak B. Berelson miał dobre zdanie o perspektywach badań językoznawczych prowadzonych tą metodą, wymieniając: „mierzenie czytelności”

(1970, 22) (w polskiej praktyce badawczej mierzenie: „zrozumiałości języka tekstu”, „trudności językowej” (Pisarek 2007; 2007a); „wykrywanie cech stylistycznych” (Berelson 1970, 24); „zastosowanie analizy ilościowej do [badań] stylu literackiego” (tamże). W tego rodzaju przedsięwzięciach jednostkami analizy byłyby całe, reprezentujące dane rodzaje czy grupy tekstów, korpusy słowne, a pomiar dotyczyłby ilości wystąpień danej cechy, tożsamej z ilością wystąpień leksemów, będących zazwyczaj nośnikami tej cechy. I tak Pisarek (2002, 199 i n.) proponuje mierzyć zrozumiałość (czy raczej niezrozumiałość) tekstu m. in. ilością wystąpień w nim wyrazów trudnych (abstrakcyjnych, długich, obcych itd.). Na tej zasadzie o natężeniu emocjonalności decydowałaby ilość wystąpień słowoform uważanych za ekspresywne, o stopniu dyrektywności — ilość wystąpień określonych modulantów i w ogóle leksemów wskazujących na przymus czy konieczność czegoś, etc.. Już prosty pomiar ilości użyć poszczególnych części mowy dawał także pojęcie o różnicach stylistycznych między tekstami, będącymi realizacjami poszczególnych gatunków dziennikarskich, a pomiary mogły być przecież bardziej skomplikowane (Pisarek 1972). Rodzaje językowych badań ilościowych omówiono w innym miejscu (Kajtoch 2008, 24 i n.).

Artykuł poniekąd przypomni te techniki, wskaże jednak jeszcze na to, że ilościowa analiza zawartości, a konkretnie jedna z jej odmian: ilościowa leksykalna analiza zawartości może mieć zastosowanie do badań nie tylko stylistycznych konwencji, ale i zdolności danego, rozumianego językoznawczo (Bartmiński/Niebrzegowska-Bartmińska 2009, 32) dyskursu do przedstawiania określonych interpretacji rzeczywistości, określonych treści.

* * *

Dyskurs to utrwalony sposób używania języka w określonych sytuacjach i grupach tekstów. Sugeruje istnienie swoistych cech rzeczywistości, do której się odnosi, ponieważ użycie pewnych środków językowych w danym celu sensowne jest, jeśli odbywa się w świecie o określonych stosunkach społecznych, cechach przyrody i cywilizacji — np. obraza: *ciemny chamie* (a nie *ciemny panie*) ma sens w społeczeństwie pamiętającym podziały klasowe, nazywanie kogoś *podstępny wężem* byłoby niemożliwe w języku Eskimosów, bo węże w lodach nie żyją.

Przedmiotem naszej porównawczej refleksji są cztery współczesne polskie dyskursy, których cech poszukaliśmy w korpusach (po 30000 słowoform) złożonych z tekstów portalu „Pudelek. Sensacja, plotka, komentarz” (www.pudelek.pl) — dalej PUD, wyboru fanzinów faszystowskich przełomu XX i XXI wieku (np. „Aryjski Opór”, „Blood&Honor”, „Ofensywa”, „Rzeźnia”, „White Empire”) — dalej FF, najważniejszego polskiego fanzina punkowego „Pasażer” (P) oraz zebranych przez M. Pyrek tekstów piosenek rockowych, wykonywanych oficjalnie lub poza cenzurą w latach osiemdziesiątych XX w. (PR).

Chodzi o korpusy tekstów typu:

We wszystkich pozytywnych opiniach podnoszony jest fakt, że udało nam się wokół antyrasistowskiego sprzeciwu zgromadzić tak różnych światopoglądowo wykonawców: anarchistów, integralnych katolików, obrazoburczych prześmiewców, rockowych poetów, punkrockowych publicystów, czołówkę niezależnych i gwiazdy rocka. [...] Natomiast niemiłą niespodzianką były głosy negujące sens takiego doboru wykonawców. [...] To dziwne jak wielu wolnościowców grzeszy nietolerancją i brakiem poszanowania dla cudzej odmienności, choć będącej często udziałem dużej części społeczeństwa, samemu jednocześnie żądając absolutnej akceptacji dla własnych nie raz ekskluzywnych pomysłów na życie. [P]

Uważam, że Edyta Górniak bardzo dobrze zrobiła, że nie zgodziła się na bycie jurorem w tym samym programie, w którym jurorem jest Doda — twierdzi. Edyta Górniak to jest zupełnie inny poziom, inna klasa, inna wiedza na temat śpiewu i nie powinna zasiadać obok Dody, która reprezentuje dużo niższy poziom zachowania, wiedzy, czy chociażby umiejętności śpiewu. Edycie Górniak nie jest potrzebne, żeby wdawać się w takie chamówki. [...] Doda na pewno... pasowałaby jej taka Edyta Górniak w jury, ponieważ Doda, żeby istnieć, potrzebuje na pewno konfliktów. Pasowałoby jej to, by się pięknie wyżywała na Edycie. Edyta aż taka za głupia nie jest. Prawdopodobnie potrafiłaby odpysknać [PUD]

Najważniejsze dla narodowego socjalisty powinno być życie, postępowanie i myślenie zgodnie z duchem aryjskim. Oznacza to życie według zasad narodowego socjalizmu — osiągnięcie własnego triumfu woli i zastosowanie tych zasad w codziennym życiu. Tak więc, musi narodowy socjalista honorowy, wierny i niezłomny w spełnianiu swych obowiązków względem narodu. Częścią tych obowiązków jest małżeństwo w obrębie własnej rasy oraz spładzanie bądź zaopiekowanie się zdrowym aryjskim potomstwem. Obowiązkiem jego jest przekonywanie innych do narodowego socjalizmu i dążenie do narodowo socjalistycznej rewolucji. [FF]

Naciskam ręką ruchomą ścianę, lecz ona wciąż trwa / Jeszcze pogładzę ją palcami i odchodzę, nie przyjdę tu sam // Od ludzi dumnych, kruchych jak sława, Ty oddzielaj nas / Od nienawiści, idei Marsa, Ty oddzielaj nas / Oooo Przed pustką ochroń nas / Przed błędnym hasłem, słowem, zasłoń nas // Będziemy wzmacniać i przesuwac Cię tak, byś naprzód szła / I podeprzemy Cię swoją siłą, gdy będziesz się chwiać / A kiedy, kiedy przyjdzie już czas, my zburzymy Cię / I zbudujemy przepiękny pałac, on jest naszym snem / Od ludzi dumnych, kruchych jak sława Ty, Ty oddzielaj nas / Od nienawiści, idei Marsa, Ty oddzielaj nas / Oooo Przed pustką zasłoń nas / Przed błędnym hasłem, słowem, Ty mocno zasłoń nas [PR]

Nietrudno obmyślić przykładowe wyrażenia, w danym dyskursie niemożliwe (poza użyciem ironicznym lub cytowaniem): *policja zasłużona dla Rzeczpospolitej* [dla P], *idiota białas* [dla FF], *istota przeżycia egzystencjalistycznego zawiera się w* [dla PUD], *dziedzinę nazywa się czasem zbiorem argumentów funkcji* [dla PR]. Natomiast w określaniu rzeczywiście istniejących cech dyskursu i stopnia ważności danego elementu czy danej cechy opisywanego dyskursu skazani jesteśmy zwykle na decyzje na poły intuicyjne i tworzenie modeli, które — jak to celnie, w odniesieniu do konstruowania językowych obrazów świata, określił R. Tokarski — „są wynikiem ekscerpcji interesujących badacza poświadczeń” (2013, 321).

Chyba że posłużymy się metodami ilościowymi, więc ustalaniem frekwencji leksykalnych elementów nie tyle nas interesujących, co powtarzających się wystarczająco często. Im częściej dany element autosemantyczny (rzeczownik, czasownik, przysłówek, przymiotnik) się powtarza, tym bardziej w korpusie tekstów się utrwała i bardziej prawdopodobne jest, że buduje trwałą rdzeń dyskursu —

panującej w korpusie językowej konwencji i sugerowanych przez nią cech świata opisywanego danym językiem.

Rozpatrując frekwencję elementów synsemantycznych, leksemów o funkcjach składniowych, modulantów itd. nie ustalimy elementów tekstowego obrazu świata [TOS], a więc:

swoistej, dokonanej w konkretnym tekście — lub w zespole tekstów — realizacji językowego obrazu świata [tj.] zbioru prawidłowości wynikłych z faktu preferowania w danym tekście lub zespole tekstów określonych konstrukcji fleksyjnych, słowotwórczych, składniowych, a przede wszystkim — określonego słownictwa, [który] wskazuje na panujący w danym tekście (zespole tekstów) pogląd na temat istnienia i funkcjonowania poszczególnych składników świata, ich związków oraz wzajemnych proporcji, a więc na takie rozumienie organizacji świata, panujących w nim hierarchii i wartości, które jest preferowane przez nadawcę danego tekstu i akceptowane przez użytkowników tegoż tekstu (Kajtoch 2008, 14).

Nie jest to korzystne, bo one z reguły zajmują pierwsze miejsca na listach rangowych, które zazwyczaj wyglądają, jak w tabeli 1.

Tabela 1. Leksemy o pierwszych 30 rangach w 4 korpusach

R	Ziny faszystowskie	Portal „Pudelek”	„Pasażer”	Piosenki rockowe
1	920 i	855 się	1125 ten/ta/to	1045 nie
2	904 w	781 w	803 być	797 i
3	667 być	656 on/ona/ono	696 i	754 być
4	614 ten/ta/to	643 ten/ta/to	611 w	570 w
5	447 z	613 być	567 z	566 ja
6	446 się	585 z	520 nie	538 się
7	412 na	576 nie	458 na	503 na
8	399 on/ona/ono	560 na	337 on	366 on
9	345 do	534 że	316 że	458 to
10	301 nie	528 i	271 a	346 ty
11	275 który	303 do	260 się	301 z
12	187 a	243 który	254 do	282 mieć/miewać
13	179 że	223 mieć	231 który	254 co, tak
14	171 o	188 jak	193 ale	249 my
15	170 walka	175 o	183 co	247 jak
16	162 rok	164 swój	181 jak, o	241 chcieć
17	160 człowiek/ludzie	157 już	154 tak, taki	238 wiedzieć
18	157 nasz	140 my	149 mieć	237 już
19	144 swój	122 bardzo [+ stopniowanie], za	135 ja	229 a
20	138 mieć	121 po	130 bardzo [+ stopniowanie]	195 ten
21	137 jak	117 gwiazda (osoba znana), rok	126 po	175 że
22	126 dla, wszystko/wszyscy/wszystek	112 duży [+ stopniowanie]	124 czy	169 wszystko/wszyscy/wszystek

R	Ziny faszystowskie	Portal „Pudelek”	„Pasażer”	Piosenki rockowe
23	121 przez	105 siebie	112 my	163 do
24	109 co	104 ale	109 płyta [muzyczna]	162 nic
25	108 za	103 mówić	108 kapela [zespół muz.]	160 nowy
26	100 my	98 chcieć	98 od	141 mój
27	99 zostać/zostawać	95 od	96 wszystko/wszyscy/wszystek, za	139 świat
28	97 po	87 czy, móc	95 nasz	138 za
29	91 rasa	85 tak	94 też	135 kto
30	88 biały (przym.)	84 a, dla	90 grać/zagrać [artystycznie]	134 tylko

Odpowiednio klasyfikując tego typu leksemy ustalić możemy jednak panujące w danym korpusie zwyczajne stylistyczne, które wskazują na pewne osobliwości zakładanego w danym dyskursie oddziaływania na odbiorcę oraz bardzo ogólne cechy procedur kreowania TOS. Badając nasze korpusy (wzięliśmy pod uwagę leksemy co najmniej trzykrotnie powtórzone) uzyskaliśmy następującą klasyfikację (tabela 2).

Tabela po pierwsze powiadamia o różnicach we frekwencji leksemów: *i*, *w*, *nie*, wskazującej na różnice w natężeniu występowania skłonności do prostego, aintelektualnego „opisu świata poprzez wyliczenie” (*i*), skłonności do opisu bardziej hierarchizującego elementy rzeczywistości i konkretniej, precyzyjniej umiejscawiającego dane fenomeny w czasie i przestrzeni (*w*) oraz tendencji do buntowniczej negacji (*nie*). Kolejne dane zgromadzone są w kolumnach.

Tabela 2: Słownictwo decydujące o cechach retorycznych, stylistycznych, światopoglądowych

Korpus	I. Dyrektywność i perswazyjność	II. Przyczynowo-skutkowość i celowość	III. Niepewność	IV. Refleksyjność
Piosenki rockowe i/w/nie – 797/570/104 my/nasz 349/60	wszystko/wszyscy/wszystek — 169; nikt — 74; nigdy — 71; każdy — 63; musieć — 35; trzeba — 29; wiele — 16; naprawdę — 16; nawet — 15; niech — 14; wreszcie — 11; byle — 8; właśnie — 8; niestety — 6; wiadomo — 6; na pewno — 4	bo — 79; by/byś — 53; więc — 27; żeby — 26; aby — 19; aż — 13; dlatego — 3	móc — 119; chyba — 13; jakby — 9; jeśli — 7; jeżeli — 7; pewnie — 3; prawie — 3	a — 229; tylko — 134; czy — 105; przecież — 27; choć — 27; lecz — 25; ale — 20; chociaż — 15; choćby — 8; jednak — 8; ani — 8; lub — 7; albo — 5; zamiast — 4; oprócz — 4; w końcu — 4;
Suma:	545	220	161	630

Korpus	I. Dyrektywność i perswazyjność	II. Przyczynowo-skutkowość i celowość	III. Niepewność	IV. Refleksyjność
Ziny faszystowskie i/w/nie — 920/904/301 my/nasz — 100/157	wszystko/ wszyscy/ wszystek — 125; wiele — 52; musieć — 41; nawet — 25; niestety — 15; jedynie — 15; należy — 15; powinien/powinna — 14; naprawdę — 11; tzw. — 11; nikt — 8; dopiero — 6; wiadomo — 5; nigdy — 23; przede wszystkim — 5; na pewno — 4; na szczęście — 3	więc — 39; aby — 34; by — 33; żeby — 21; bo — 20; dlatego — 17; gdyż — 17; ponieważ — 14; dzięki [komu; czemu] — 9; aż — 8; na skutek — 8; z powodu — 6; w imię — 6; zatem — 5; na mocy — 3; na rzecz — 4;	móc — 78; jeżeli — 23; może — 17; około — 17; może — 17; jeśli — 12; według — 9; prawie — 7; chyba — 4; jakby — 4; na razie — 4; z drugiej strony — 5	a — 187; ale — 80; czy — 73; jednak — 32; lub — 32; na przykład — 21; albo — 14; lecz — 11; w końcu — 11; zaś — 11; pomimo — 10; ani — 8; czyli — 8; oprócz — 8; choć — 6; mimo to — 6; bądź — 5; chociaż — 5; przecież — 5; choćby — 4; z resztą — 4; zwłaszcza — 4; ani [...] ani — 3; jednakże — 3
Suma:	378	245	197	551
„Pasażer” i/w/nie — 696/611/520 my/nasz — 112/95	musieć — 25; wszystko/ wszyscy/ wszystek — 96; każdy — 51; nawet — 44; wiele — 34; trzeba — 27; właśnie — 22; naprawdę — 20; nigdy — 19; oczywiście — 16; na pewno — 16; powinien/powinna — 15; zawsze — 12; nikt — 12; dopiero — 11; wszelki — 11; niech — 10; na szczęście — 8; tzw. — 7; jedynie — 5; przede wszystkim — 5; wreszcie — 5; wyłącznie — 5	bo — 81; by — 58; więc — 37; żeby — 18; aby — 15; aż — 15; gdyż — 14; dlatego — 9; ponieważ — 8; dzięki [komu; czemu] — 7; skoro [...] to — 4; z racji — 3	móc — 52; może — 43; chyba — 39; jeśli — 38; jeżeli — 19; jakby — 16; zapewne — 16; prawie — 12; na razie — 6; podobno — 6; według — 6; pewnie [tc. zapewne] — 5; z drugiej strony — 4; ok. [około] — 4; mniej więcej — 3	a — 271; ale — 193; czy — 124; tylko — 72; jednak — 54; niestety — 40; choć — 38; lub — 36; albo — 33; czyli — 28; ani — 23; zresztą — 20; zwłaszcza — 17; poza tym — 15; za to — 15; zbyt — 15; przecież — 15; mimo — 14; wbrew — 10; oprócz — 10; chociaż — 9; natomiast — 9; przynajmniej — 9; w dodatku — 9; mimo że — 7; w końcu — 7; lecz — 6; tymczasem — 6; zamiast — 6; choćby — 5; tylko że — 5; z kolei — 5; w ramach — 5; swoją drogą — 4; nota bene — 4; acz — 3; mimo to — 3; z arównno [...] jak i — 3; zaś — 6
Suma:	476	269	269	1154
Portal „Pudelek” i/w/nie — 528/781/576 my/nasz — 140/26	wszystko/wszystek — 78; nawet — 58; musieć — 50; właśnie — 47; wiele — 39; naprawdę — 38; każdy — 33; oczywiście — 30; powinien/powinna — 28; nigdy — 26; zawsze — 24; niestety — 22; wiadomo — 21; nikt — 19; na pewno — 15; dopiero — 15;	żeby — 59; by — 36; więc — 31; bo — 30; aż — 23; dzięki [komu; czemu] — 17; aby — 15; dlatego — 13; z powodu — 9; ponieważ — 6; skoro — 6; gdyż — 5; z racji — 3	móc — 87; może — 37; chyba — 32; podobno — 24; jeśli — 21; jeżeli — 9; na razie — 8; prawie — 7; według — 7; pewnie [tc. zapewne] — 7; jakby — 6; około — 4	ale — 114; czy — 87; a — 83; jednak — 77; tylko — 74; niestety — 22; w końcu — 14; zbyt — 12; za to — 8; natomiast — 8; chociaż — 7; albo — 7; ani — 7; pomimo — 7; czyli — 7; mimo — 5; przecież — 5; choć — 5; zamiast — 5; lub — 4; w dodatku — 4; mimo że — 4; zarówno [...]

Korpus	I. Dyrektywność i perswazyjność	II. Przyczynowo-skutkowość i celowość	III. Niepewność	IV. Refleksyjność
	jedynie — 15; trzeba — 10; przede wszystkim — 9; wreszcie — 7; na szczęście — 6; niech — 5; wszelki — 3; wyłącznie — 3			jak i — 4; tymczasem — 4; tylko że — 4; swoją drogą — 3; zresztą — 3; zwłaszcza — 3; poza tym — 3; oprócz — 3; przynajmniej — 3; na przykład — 3
Suma:	601	253	249	599

Kolumna I gromadzi słownictwo od dawna badane przez prasoznawców (Pisarek 1976). Wiadomo że leksemy typu *wszyscy, każdy, zawsze, nikt, nigdy, nigdzie* tworzą „wielki kwantyfikatory” znany w retoryce (Pszczółowski 1963, 106-108); *musieć, trzeba, powinien* — świadczą o dyrektywności, przedstawianiu danego stanu jako koniecznego; *naprawdę, oczywiście* — wskazują na próby przekonania czytelnika o prawdziwości głoszonego; a *niestety, na szczęście* itd. wprowadzają emocjonalną, ważną perswazyjnie ocenę.

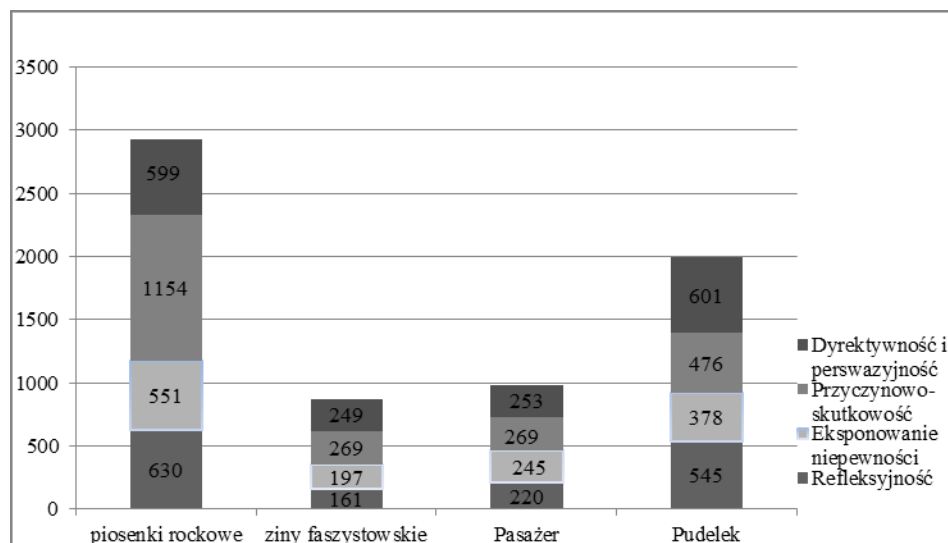
Kolumna II gromadzi słownictwo, którego używa się, chcąc przedstawić świat jako spójny, gdzie każda przyczyna ma swój skutek, możliwe jest połączenie faktów w ciągi przyczynowo-skutkowe, aby je można było zrelacjonować w narracji.

Kolumna III wskazuje z kolei na zdolność wyrażania wątpliwości, relacjonowania czegoś jako tylko prawdopodobnego (Dobek-Ostrowska/Fras/Ociepka 1997, 94-95) itd.

Kolumna IV gromadzi cząstki używane przy konstruowaniu zdań złożonych, zwłaszcza przyzwolenia (*mimo to, chociaż, choć*), konstrukcji wskazujących na istnienie alternatywnych możliwości (np. *lub, albo*), czy pewnych innych złożonych stosunków między opisywanymi faktami, stanami, przedmiotami (*zamiast, zwłaszcza, z kolei, na przykład*). Wszystkie te cząstki charakterystyczne są dla wypowiedzi o intelektualnym charakterze, starających się oświetlać różne aspekty spraw, być dokładnymi.

Porównanie ich frekwencji w 4 korpusach przyniosł wykres 1. Możemy wnioskować z niego parę niespodziewanych ustaleń. Np. że najbardziej intelektualny charakter ma dyskurs punkowy, dwa razy mniej od niego intelektualne są piosenki rockowe, a zdecydowanie najmniej intelektualne są teksty korpusu FF. Także korpusowi P należy się pierwszeństwo, jeśli chodzi o eksponowanie niepewności i umiejętność opowiadania historii, o co raczej podejrzewalibyśmy dyskurs „Pudelka”. PUD jest też zdecydowanie najbardziej perswazyjny, o co raczej podejrzewać moglibyśmy zideologizowany dyskurs korpusu FF. Rozważania można kontynuować. A gdyby jeszcze kategorie I-IV uzupełnić o podkategorie, gromadzące słownictwo służące do wzmocnienia poszczególnych aspektów refleksyjności, dyrektywności, teleologiczności itd. — można by znakomicie pomniejszyć ogólność rozważań, na co w tym krótkim tekście miejsca nie ma.

Wykres 1: Cztery korpusy — ich cechy stylistyczno-ideologiczne na podstawie frekwencji wybranego słownictwa (wartości liczbowe — patrz także rubryki „Suma” w tabeli 2)



Elementy TOS związanego z dyskursami swoistymi dla tekstów 4 korpusów, ich ważność i powiązanie można wykryć narzędziami subtelniejszymi: 1) listami rangowymi rzeczowników powtarzających się w korpusach; 2) schematami tematycznych podziałów rzeczowników tworzących korpusy. Uwzględniliśmy w nich leksemy o występowaniu przynajmniej czterokrotnym, więc już wydatnie służące tworzeniu konwencji użycia języka (dyskursu).

Wstępne wnioski można wyciągnąć z list frekwencyjnych. Tak się składa, że rzut oka już na pierwszych 5 miejsc takiej listy (patrz tabela 3) należycie informuje, o co chodzi w dyskursie — a zatem i TOS — danego korpusu.

Tabela 3. Rzeczowniki o pierwszych 25 rangach w 4 korpusach

R	Ziny faszystowskie	Portal „Pudelek”	„Pasażer”	Piosenki rockowe
1	170 walka	117 rok	109 płyta [muzyczna]	139 świat, ludzie/człowiek
2	162 rok	78 wszystko/wszystek	108 kapela [zespół]	120 głowa
3	160 człowiek/ludzie	65 film	89 zespół [muz.]	89 noc
4	126 wszyscy/wszystek	59 aktorka, dziecko	87 rok	80 twarz
5	91 rasa	52 czas, życie	74 człowiek	72 krew, 72 życie

R	Ziny faszystowskie	Portal „Pudelek”	„Pasażer”	Piosenki rockowe
6	76 naród, życie	45 scena, związek	66 koncert	70 strona, wojna
7	70 skin/skinhead	40 aktor, media	57 strona [płyty, kasety, czasopisma]	63 — czas, rok
8	67 grupa, ruch (organizacja)	39 rola (postać), zdjęcie	51 rzecz	60 znak
9	65 członek (np. organizacji)	37 dzień, mąż, osoba	50 czas (ciąg chwil), kawałek [utwór muz.]	58 słowo
10	59 zespół	35 wywiad	45 muzyka	57 sen
11	55 świat	34 tabloid	42 tekst/text, scena [środowisko]	52 edukacja
12	54 narodowy socjalizm	31 kobieta, miejsce, piosenkarka	38 sytuacja	51 ręka
13	53 czas (ciąg chwil)	30 dom, program, sprawa	37 kasetka [magnetofonowa i video]	49 fabryka
14	46 wojna	29 plan	35 wywiad	48 śmierć
15	44 kraj	28 koncert, taniec, tydzień, występ	33 numer, sprawa	46 Bóg, fala
16	43 organizacja	27 kariera	31 lp [long play]	45 dom, oko
17	42 koncert	26 człowiek/ludzie, miesiąc, dziennikarz, odcinek (programu), słowo, żona	30 zin [odmienia się]	44 zmiana
18	40 lud (zbiorowość)	25 informator, rozmowa	29 życie	43 kopulacja
19	37 historia	24 kochanka	28 punk/pank [styl muz.]	39 prawo, strach
20	36 sposób	23 show, ślub	27 typ [rodzaj]	38 dłoń
21	34 rasizm, sprawa	22 para (o dwojce osób), seks, powód (czegoś), show-biznes	26 temat	37 tłum
22	31 pismo (gazeta), praca, prawo	21 pieniądz, rodzina, romans	25 część	36 myśl, slogan
23	30 cywilizacja	20 informacja, świat	23 pieniądz, początek, singiel [rodzaj płyty]	34 dżuma, polityk
24	28 działalność, miejsce, pieniądz	19 ojciec, producent, raz, widz, sytuacja	22 inny, miejsce, numer [utwór muz.], wokal [głos], państwo [instytucja], świat	33 niebo
25	26 muzyka, religia	18 impreza, płyta, praca, strona, przyjaciel, tysiąc (rzecz.)	21 kolega, punk/pank/punx [członek sukultury], sposób	32 mur

Jak widać, w FF chodzi przede wszystkim o *walkę i rasę*, w PUD o *filmy, aktorki i nowości (rok, czas)*, w P o *płyty, kapele, zespoły*, bo „Pasażer” to jednak czasopismo muzyczne, a TOS korpusu PR zorganizowany był wokół egzystencjalnych niepokojów jednostki i zbiorowości (*człowiek/ludzie, świat, głowa, twarz, noc*).

Ponieważ w badaniach prowadzonych za pomocą list rangowych zakłada się, że im miejsce leksemu wyżej, tym jest on ważniejszy (a więc jego denotat ważniejszy jest w TOS tekstów korpusu) można listy frekwencyjne czytać trochę inaczej. Np. ustalić, że w świecie wyznaczanym przez konwencje FF rasa ważniejsza jest od narodu (*rasa* — ranga 5, *naród* — r. 6), w PUD — związek od domu (*związek* — r. 6, *dom* — r. 13), w P — płyta od muzyki w stylu punk (*płyta* — r. 1, *punk* — r. 18), w PR — człowiek od Boga (*człowiek/ludzie* — r. 1, *Bóg* — r. 15). Albo inaczej, można np. wyczytać, że dla faszysty naród jest tak samo ważny jak koncert dla punka (*naród* — r. 6 w FF; *koncert* — r. 6 w P.).

Precyzyjniejszym narzędziem pozwalającym ustalić zarysy TOS swojego dla tekstów danego korpusu (a zatem — pośrednio — ustalić cechy dyskursu, który buduje TOS o właśnie takich zarysach) jest specyficznie wykorzystany podział tematyczny rzeczowników korpusu. Zbudowaliśmy taki, opierając się z grubsza na tradycji leksykograficznej Rudolfa Halliga i Walthera von Wartburga oraz Franza Dornseiffa (Wierzchowski 1955; Markowski 1990; Kajtoch 2008, 221-343) i otrzymaliśmy następujące wyniki (tabela 3).

Tabela 3. Rozkład leksemów 4 korpusów w polach tematycznych

	PUDELEK		ZINY FASZYSTOWSKIE	
	LEKSEMY	SŁOWOFORMY	LEKSEMY	SŁOWOFORMY
SFERA 1. ONTOLOGICZNE I FIZYCZNE WŁASNOŚCI ŚWIATA	53	746 = 17,2%	105	1244 = 22,7%
I. Istnienie	11	155	41	486
	rzecz, obecność, wszystko, koniec, brak		stworzenie, rzecz, istnienie, proces, ewolucja, rozwój	
II. Czas	23	395	33	488
	czas, pora, chwila, rok, dzień, moment		czas, okres, moment, historia, przeszłość, tradycja, przetrwanie	
III. Przestrzeń	4	62	15	131
	miejsce, kierunek, środek		góra, krąg, granica, miejsce, orientacja, wschód, południe	
IV. Jakości ilościowe (zbiory danych elementów)	7	78	11	86
	para, tysiąc, połowa, większość		numer, liczba, reszta, połowa	
V. Dane ilościowe (poszczególne elementy)	2	8	0	0
	kilo, waga		BRAK	
VI. Kolejność	3	22	1	5
	seria, ranking		szereg	
VII. Jakości charakterystyczne przede wszystkim dla ruchów i procesów	2	11	3	44
	poziom, siła		siła, potęga	
VIII. Ruch uniwersalny, możliwy dla każdego obiektu	1	15	1	4
	powrót		powrót	

SFERA 2. NATURA	40	428 = 9,9%	28	438 = 8%
IX. Przyroda nieożywiona	4	36	12	146
	świat, lód, ziemia		świat, natura, ogień, ziemia, morze, planeta	
X. Przyroda ożywiona	36	392	16	292
	dziecko, kobieta, chłopak, głowa, oko, włos, życie, ciąża, seks, pocałunek, krok, gest		młodzież, dziecko, rasa, biały, murzyn, krew, ręka, serce, życie śmierć	
SFERA 3. OSOBA	53	467 = 10,8%	73	609 = 11,1%
XI. Zmysły ludzkie	1	5	1	4
	słuch		kolor	
XII. Emocje	8	72	1	5
	kochanka, miłość, uczucie, żaloba		nienawiść	
XIII. Stałe tendencje emocjonalne	7	56	10	79
	przyjaciel, żal, szkoda		stan, postawa, nadzieja, duma, dobro, zło	
XIV. Wola	7	51	10	89
	decyzja, wpływ, warunek, wybór		wpływ, jedność, wola, wolność	
XV. Myśl i mowa	30	283	51	432
	fakt, prawda, tajemnica, doświadczenie, propozycja, słowo, rozmowa		prawda, fakt, temat, zainteresowanie, idea, wartość, doktryna, słowo, nazwa	
SFERA 4. TWÓRCZOŚĆ	72	908 = 20,9%	32	367 = 6,7%
XVI. Kultura artystyczna i piśmiennictwo	scena, występ, widz, piosenkarka, koncert, show, zdjęcie		scena, artykuł, zespół, koncert, muzyka, płyta, sztuka, czytelnik	
SFERA 5. LOS I CZYN	34	294	45	400
XVII. Bieg życia i działania ludzkiego	sprawa, sytuacja, sukces, okazja, wydarzenie		problem, sytuacja, szansa, sprawa, działalność, sposób, cel	
SFERA 6. SPOŁECZEŃSTWO	156	1499 = 34,5%	190	2428 = 44,3%
XVIII. Cywilizacja życia codziennego	17	88	5	31
	sukienka, strój, but, lekarz, mieszkanie, miasto		naszywka, ulica, miasto	
XIX. Mniej zinstytucjonalizowane sfery aktywności	56	566	47	515
	skandal, walka, narkotyki, impreza, klub, piłkarz, samochód, media, wywiad, tabloid		walka, wróg, zagrożenie, impreza, droga, pismo, dzieło, media	
XX. Własność	26	218	10	123
	nagroda, kariera, praca, pieniądze, złoty (waluta), dolar, model, modelka		interes, podatek, praca, współpraca, pieniądze, dolar, środki	
XXI. Człowiek pracujący i produkujący	4	48	2	8
	producent, projekt, firma		pojazd, produkcja	
XXII. Instytucje społeczne	53	579	126	1751
	prezydent, państwo, policja, kościół, ksiądz, związek, mąż, żona, ślub, rodzina, styl		naród, narodowy socjalizm, kraj, lud, wojna, prawo, więzienie, religia, chrześcijaństwo, lud, ruch	
Suma	408	4342 [=100%]	473	5486 [=100%]

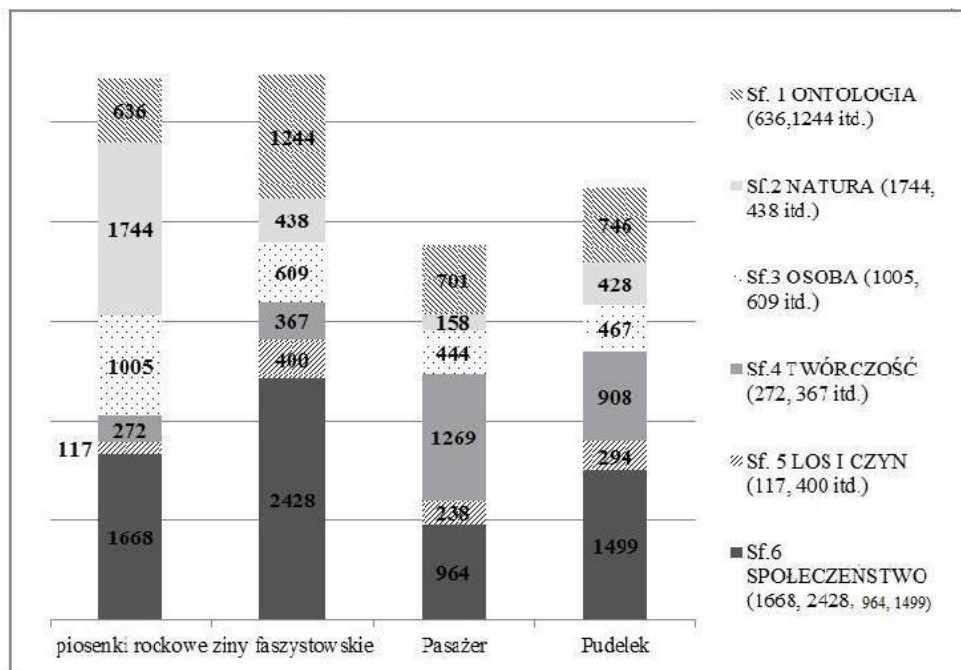
	PASAŻER		PIOSENKI ROCKOWE	
	LEKSEMY	SŁOWOFORMY	LEKSEMY	SŁOWOFORMY
SFERA 1. ONTOLOGICZNE I FIZYCZNE WŁASNOŚCI ŚWIATA	61	701 = 18,6%	40	636 = 11,7%
I. Istnienie	23	251	9	80
	rzecz, rzeczywistość, obecność, składnik, istota, forma		struktura, część rodzaj, system, układ, koniec, zero, zanik	
II. Czas	18	274	12	333
	odcinek, pora, czas, chwila, rok, tempo, korzenie		czas, chwila, noc, rok, era, godzina, przyszłość, zmiana	
III. Przestrzeń	8	59	15	191
	przestrzeń, krąg, granica		strona, miejsce, bok (kierunek), dół (kierunek), granica	
IV. Jakości ilościowe (zbiory danych elementów)	9	100	1	10
	numer, reszta, garść		numer	
V. Dane ilościowe (poszczególne elementy)	0	0	0	0
	BRAK		BRAK	
VI. Kolejność	1	5		
	seria		BRAK	
VII. Jakości charakterystyczne przede wszystkim dla ruchów i procesów	1	6	2	18
	stopień		siła, moc	
VIII. Ruch uniwersalny, możliwy dla każdego obiektu	1	6	1	4
	ruch		pęd	
SFERA 2. NATURA	18	158 = 4,2%	101	1744 = 32%
IX. Przyroda nieożywiona	3	38	34	501
	świat, energia, woda		świat, fala, gaz, niebo, wiatr, powietrze, ziemia (obszar), woda, ogień, słońce	
X. Przyroda ożywiona	15	120	67	1243
	dziecko, pan, pani, chłopak, życie, ręka, serce		drzewo, róża, dziecko, pan, kobieta, chłopiec, pies, małpa, krew, życie, śmierć	
SFERA 3. OSOBA	62	444 = 11,3%	92	1005 = 18,5%
XI. Zmysły ludzkie	4	23	19	155
	wizja, spojrzenie		głos (dźwięk), cisza, cień, światło, ciemność	
XII. Emocje	4	22	11	121
	śmiech, emocja, miłość przyjemność, ból		nienawiść, ból, płacz, uśmiech, uczucie	
XIII. Stałe tendencje emocjonalne	6	34	17	212
	stan, instynkt, nadzieja, honor		strach, nadzieja, psychopata, zdrada, zło, wina, lęk, przyjaciel, litość	
XIV. Wola	6	61	15	125
	racja, wpływ, wola, kontrola		marzenie, wolność, niewola, plan, władza	
XV. Myśl i mowa	42	303	30	392

	temat, fakt, sens, pomysł, doświadczenie, idea, zasada, słowo	słowo, wiadomość, znak, zdanie (pogląd), myśl, rozum, głupiec
SFERA 4. TWÓRCZOŚĆ	92 1269 = 33,6%	26 272 = 5%
XVI. Kultura artystyczna i piśmiennictwo	scena (środowisko), tekst, okładka, koncert, muzyka, płyta, film, wideo, fotka, zdjęcie, komiks	ohyda, sztuka slogan, kurtyna, film, teatr, manekin
SFERA 5. LOS I CZYN	23 238	13 117
XVII. Bieg życia i działania ludzkiego	sytuacja, przypadek, sprawa, działalność, początek, sposób	czyn, problem, los (bieg wydarzeń), cel, cena [niewyrażona pieniądzem]
SFERA 6. SPOŁECZEŃSTWO	99 964 = 25,5%	151 1668 = 30,8%
XVIII. Cywilizacja życia codziennego	5 39	35 352
	cukier, przepis, miasto, śmieć (o wyglądzie)	dom, mur, ulica, drzwi, ściana, miasto, śmieć, pralnia
XIX. Mniej zinstytucjonalizowane sfery aktywności	29 329	31 250
	wino, alkohol, marihuana, impreza, otoczenie, wywiad, zin, media	parada, telefon, gazeta, radio, pokój (zgoda), droga, wróg, bomba, broń
XX. Własność	17 155	9 56
	rynek, koszt, podatek, emerytura, praca, robota, pieniądz, złoty (waluta)	pieniądz, konsument, handlarz, kram, praca
XXI. Człowiek pracujący i produkujący	6 42	15 151
	firma, produkcja, produkt	fabryka, automat, maszyna, zakład (pracy)
XXII. Instytucje społeczne	42 399	61 859
	państwo, kraj, kampania, biblia, urzędnik, instytucja, punk, załoga, kolega, ekipa, moda	kultura [w znaczeniu socjologicznym], cywilizacja, edukacja, szkoła, wojna, prawo, rozkaz, karabin, kałasznikow, ojczyzna
Suma	355 3774 [=100%]	423 5442 [=100%]

Sam podział, dzięki któremu mogliśmy zaliczyć uporczywie (minimum czterokrotnie) się powtarzające rzeczowniki w ramy poszczególnych sfer i kategorii tematycznych — jak widać — jest uniwersalny dla czterech korpusów i pewnie jeszcze wielu innych tekstów, więc interesujących nas dyskursów i TOS-ów nie różnicuje. Można także go użyć do klasyfikowania wszystkich części mowy i to występujących częściej lub rzadziej. W momencie jednak, gdy policzyliśmy frekwencję leksemów pogrupowanych w dane sfery i kategorie — otrzymaliśmy informację na temat intensywności, z jaką teksty danych korpusów korzystają z określonych możliwości tematycznych (dokładniej — z rzeczowników o takich a nie innych znaczeniach, wchodzących w zakres danej tematycznej sfery lub kategorii). Ta informacja pozwoliła korpusy (tj. TOS-y ich tekstów) bardzo wyraziście różnicować. Może nie widać tego dokładnie w tabeli 3 (pokazuje tylko przykłady słownictwa o największych frekwencjach w grupie, choć uwzględnia frekwencje łączną dla wszystkich leksemów), ale wykresy — niestety tylko ogólnie wskazujące na użytkowanie sfer bez wgłębiania się w kategorie — nieźle wyniki obrazują.

Wykres 2 pokazuje wyniki obliczeń frekwencji interesującego nas słownictwa zapisane w liczbach bezwzględnych.

Wykres 2. Zestawienie w liczbach bezwzględnych intensywności wykorzystywania słownictwa tworzącego 6 sfer tematycznych w 4 korpusach tekstów

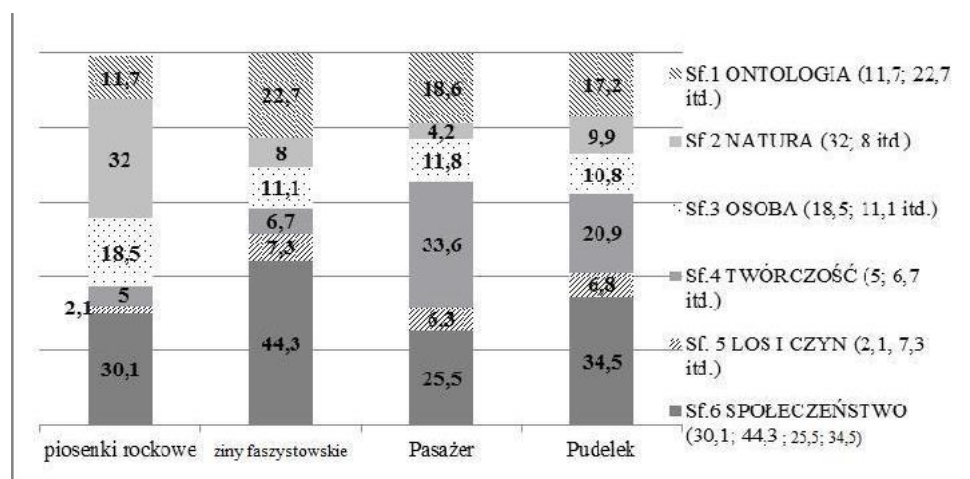


Główna informacja, którą możemy tu wyczytać pokazuje, że dane korpusy (równe, liczące po 30000 słowoform) budują konwencję w sposób bardziej lub mniej intensywny. Najbardziej skonwencjonalizowany jest korpus faszystowski (najwięcej tam rzeczowników przynajmniej czterokrotnie powtarzających się), najmniej — punkowy, bo tam rzeczowników o mniejszych frekwencjach jest przypuszczalnie najwięcej.

Różnice między frekwencjami występują też dzięki różnicom w ilości nazw własnych. Obliczyliśmy, że w korpusie PR wystąpiło 157 słowoform nazw własnych, w korpusie FF — 1116, w korpusie PUD — 387, w korpusie P — 597. Wyraźnie to świadczy, o — że tak się wyrazimy — stopniu kontaktu tekstów korpusu z rzeczywistością realną (najmniejsza — w tekstach piosenek). Ważne mogą być też całkiem inne sfery tej realnej rzeczywistości (w FF — jest 446 słowoform nazw związanych z polityką, w PUD — 198 słowoform nazw związanych z mediami).

No i nawet te same sfery rzeczywistości mogą być widziane zupełnie różnie, mieć inaczej rozłożone akcenty. W dyskursie faszystowskim nazwy miejscowe o największej frekwencji to: *Polska* (57), *Europa* (51) *Niemcy* (19), punkowym: *Polska* (22), *Japonia* (11), *Anglia* (9), w „Pudelku”: *Polska* (23), *Warszawa* (16), *Europa* (11), polskich piosenek rockowych lat osiemdziesiątych — *Polska* (26), *Babilon* (15), *Moskwa* (9).

Wykres 3. Procentowe zestawienie intensywności wykorzystywania słownictwa tworzącego 6 sfer tematycznych w 4 korpusach tekstów



Natomiast wykres 3 pozwala ogólnie ocenić, na jaką sferę ludzkiej działalności dyskurs danego korpusu był nakierowany przede wszystkim lub marginalnie, w co najsilniej zaangażowane były te środki leksykalne, które już z całą pewnością (tzn. jako wysoce powtarzalne) tworzyły konwencję. Okazuje się, że TOS zinów faszystowskich rozwinął obrazy życia społecznego, w sposób nikły angażując się w tematykę egzystencjalną, dyskurs piosenek rockowych pragnął widzieć człowieka na tle utopijnych wizji natury, język punkowego „Pasażera” przystosowany był do omawiania problematyki artystycznej, Pudelek natomiast plotkuje o „gwiazdach”, ani nie opisując zbyt intensywnie ich działań artystycznych, ani nie zwracając za bardzo uwagi na ich walory osobowościowe. Wszystkie te tendencje występowały zaś w precyzyjnie wyliczalnym natężeniu.

Jak sądzimy, nietrudno wyobrazić sobie bardziej szczegółowe, procentowe wyliczenia, dotyczące nie tylko tematycznych sfer, ale i poszczególnych kategorii i podkategorii, a więc — tym samym — dalsze i konkretniejsze perspektywy leksykalnej analizy zawartości zastosowanej do badań dyskursu (a więc i TOS) ujawniającego się w... czy charakterystycznego dla... różnorodnych korpusów tekstowych.

Ze względów zrozumiałych nie możemy w krótkim artykule tych wyliczeń dokonać, tak samo jak sięgnąć do kolejnych, stosowanych w leksykalnej analizie zawartości ilościowych metod badawczych — na przykład dokonać pomiaru aksjologicznego (Kajtoch 2008, 127 i n.; Rodak 2008, 169 i n.), którego celem byłoby wskazanie różnic w nasyceniu poszczególnych korpusów tekstowych słownictwem odnoszącym się do poszczególnych wartości (społecznych, witalnych, hedonistycznych itd.). Nie zaprezentujemy także wyliczeń dotyczących frekwencji słów kluczowych (zob. Wyka 1968) lub też słów sztandarowych (zob. Pisarek 2002a), odsyłając czytelników do przygotowywanej przez nas większej publikacji.

Bibliografia

- BARTMIŃSKI, J. / NIEBRZEGOWSKA-BARTMIŃSKA, S. (2009), *Tekstologia*. Warszawa.
- BERELSON B. (1952), *Content Analysis in Communication Research*. New York.
- (1970) *Analiza treści*. W: anonim (red.), *Teoria i metodologia badań nad prasą za granicą (prze-kłady)*. Kraków, 5-62.
- DOBIEK-OSTROWSKA, B./FRAS, J./OCIEPKA, B. (1997), *Teoria i praktyka propagandy*. Wrocław.
- JEZYK, A. (2005), *Najmłodsza polska poezja w świetle analizy zawartości*. W: *Zeszyty Prasoznaw-cze*. 3-4, 100-116.
- KAJTOCH, W. (2008), *Językowe obrazy świata i człowieka w prasie młodzieżowej i alternatywnej*, t.1. Kraków.
- KIKLEWICZ, A. (2010), *Tęcza nad potokiem. Kategorie lingwistyki komunikacyjnej, socjolingwisty-ki i hermeneutyki lingwistycznej w ujęciu systemowym*. Łask.
- LISOWSKA-MAGDZIARZ, M. (2006), *Analiza tekstu w dyskursie medialnym. Przewodnik dla studen-tów*. Kraków.
- MARKOWSKI, A. (1990), *Leksyka wspólna różnym odmianom polszczyzny, I-II*. Warszawa.
- MYTNIK, O. (2007) *Jak krytycy Ruchu Muzycznego oceniają muzykę, jej wydania i wykonanie*. W: *Zeszyty Prasoznawcze*. 3-4, 141-158.
- PISAREK, W. (1972), *Frekwencja wyrazów w prasie. Wiadomości — komentarze — reportaże*. Kraków.
- (1976), *Język służy propagandzie*. Kraków.
- (1983), *Analiza zawartości prasy*. Kraków.
- (2002) *Nowa retoryka dziennikarska*. Kraków.
- (2002a) *Polskie słowa sztandarowe i ich publiczność*. Kraków.
- (2007) *Recepty na zrozumiałość wypowiedzi*. W: *tenże, O mediach i języku*. Kraków 2007, 223-244.
- (2007a) *Jak mierzyć zrozumiałość tekstu?* W: *tenże, O mediach i języku*. Kraków, 245-262.
- (red.) (2006), *Słownik terminologii medialnej*. Kraków.
- PSZCZOŁOWSKI, T. (1963), *Umiejętność przekonywania i dyskusji*. Warszawa.
- PYREK, M., *Tematy, emocje, wartości w tekstach polskich piosenek rockowych z lat 1980-1989* [praca niewydana].
- RODAK, M. (2007), *Polskie i niemieckie Bravo — porównanie*. W: *Zeszyty Prasoznawcze*. 3-4, 159-175.
- TOKARSKI, R. (2013), *Światy za słowami. Wykłady z semantyki leksykalnej*. Lublin.
- WIERZCHOWSKI, J. (1955), *Dwie koncepcje słowników niealfabetycznych*. W: *Poradnik Językowy*. 4, 127-133.
- WYKA, K. (1968), *O potrzebie historii literatury*. Kraków.

Prospects for lexical content analysis

The article shows how, on the basis of four short and distinct text corpuses, quantitative analysis of the vocabulary may lead to the development of stylistic profile and outlines the so called textual worldviews (linguistic worldviews) specific to the texts generated within the specific discourses. The authors have focused in this article primarily on vocabulary excerpted from the gossip website *Pudelek*, fascist zines, punk zines and Polish rock songs from the 80s.