

NATALIA ŚMIEŁOWSKA

ORCID: 0000-0001-5929-0792

Uniwersytet Wrocławski, Polska

KAMIL WABNIC

ORCID: 0000-0001-9762-3420

Uniwersytet Wrocławski, Polska

Rozmowa egzaminacyjna – wybrane cechy stylu wypowiedzi egzaminatora i egzaminowanego. Analiza korpusowa

Wprowadzenie

Celem artykułu jest porównanie wybranych cech stylu wypowiedzi osoby egzaminującej oraz osoby egzaminowanej w sytuacji komunikacyjnej, jaką jest rozmowa egzaminacyjna. Badany materiał pochodzi z korpusu zebranego w ramach niemiecko-brytyjsko-polskiego projektu badania mówionej odmiany języka naukowego GeWiss (2019). W badaniu wykorzystano metodę komparatywną. Materiał badawczy zawiera dwa zbiory tekstów w równorzędnej relacji, a podstawę porównania tworzą wybrane cechy stylistyczne porównywanych wypowiedzi, między innymi frekwencja rzeczowników i czasowników czy stosowane konstrukcje syntaktyczne.

1. Rozmowa egzaminacyjna

Rozmowa egzaminacyjna to jedna z odmian gatunku mowy, jakim jest rozmowa (Majewska-Tworek 2014:25)¹. Aleksander Wilkoń zaproponował rozróżnienie na trzy typy gatunków mowy: gatunki mowy naturalnej (potocznej), gatunki mowy

¹ Termin „rozmowa egzaminacyjna” został zaczerpnięty z pracy „Niepłynność wypowiedzi w oficjalnej odmianie polszczyzny. Propozycja typologii” Anny Majewskiej-Tworek.

starannej, kulturalnej, tworzącej kod rozwinięty, oraz gatunki mowy oficjalnej, publicznej. Rozmowa została przypisana do gatunków mowy naturalnej (potocznej), jednak badacz podkreśla, że istnieją sytuacje, w których dany typ wypowiedzi może występować we wszystkich odmianach, oraz zauważa, że granice gatunków mówionych są niewyraźne, co skłania do traktowania z dozą sceptycyzmu próby ich arbitralnego wyodrębnienia (Wilkoń 2003:51–52). Urszula Żydek-Bednarczuk zwraca uwagę na to, że rozmowa jest podporządkowana działalności językowej. Badaczka wskazuje, że „różnica między dialogiem a rozmową polega na tym, że rozmowa – będąca pojęciem węższym – występuje w odmianie mówionej, kontakcie bezpośrednim w odmianie oficjalnej i nieoficjalnej, dialog zaś może wystąpić zarówno w odmianie pisanej, jak i mówionej” (Żydek-Bednarczuk 1994:30–31). Anna Majewska-Tworek zauważa, że we współczesnej polszczyźnie znajdują się nazwy typów rozmów, które mają oficjalny charakter – jako przykłady można przywołać między innymi rozmowę kwalifikacyjną, rozmowę rekrutacyjną czy rozmowę wstępną. W codziennej polszczyźnie nie funkcjonuje wyrażenie „rozmowa egzaminacyjna”, użytkownicy stosują słowo „egzamin” (lub z uszczegółowieniem „egzamin ustny”) (Majewska-Tworek 2014:27). Jednakże, jak zauważa Majewska-Tworek, egzamin ustny przyjmuje formę rozmowy o swoistym charakterze interakcyjnym.

Sytuacja komunikacyjna egzaminu ustnego wpisuje się w typologię rozmowy, którą wyodrębniła Bożena Taras. W typologii tej zostały uwzględnione następujące elementy: sposób przekazu informacji (słowny, pozasłowny), tworzywo wypowiedzi (mowa, pismo), ramy czasoprzestrzenne (jedność czasu i przestrzeni lub jej brak), kanał wymiany informacji (wzrokowy, słuchowy, dotykowy) oraz kontakt odbiorców (oficjalny, nieoficjalny) (Taras 2003:42–43). Uczestnicy rozmowy egzaminacyjnej przekazują informacje werbalnie. Ich kontakt jest bezpośredni, *face to face*, i ma charakter oficjalny. Zatem egzamin ustny – według rozróżnienia Taras – można uznać za rozmowę. Warto dodać, że Stanisław Gajda zaliczył egzamin oraz rozmowę naukową do gatunków czy odmian gatunkowych², które pochodzą od rozmowy potocznej (Gajda 1982:174). Jest to kolejny przykład powiązania egzaminu ustnego z rozmową.

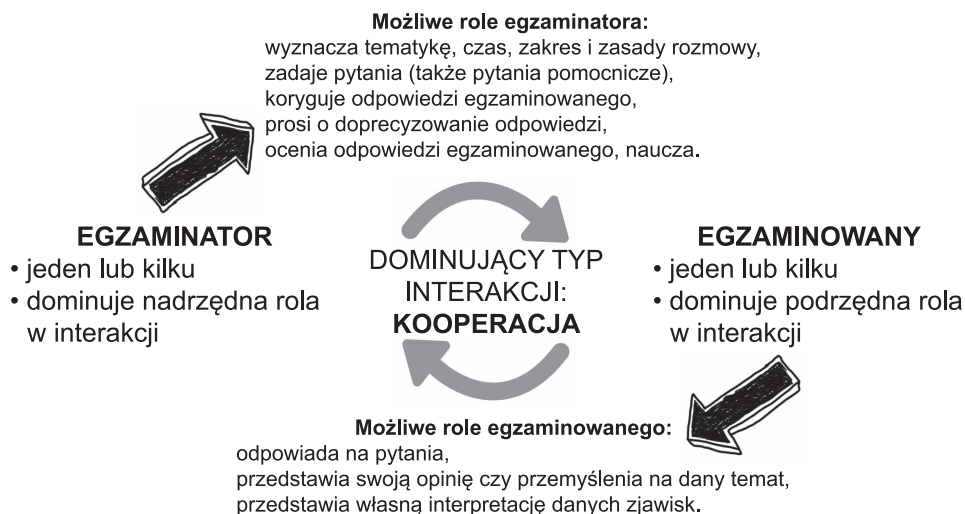
Cechy rozmowy egzaminacyjnej opisała Majewska-Tworek³. Ważne wydają się dwa elementy – rola uczestników rozmowy i jej konsekwencje w interakcji oraz typ relacji między uczestnikami rozmowy egzaminacyjnej. Badaczka przyporządkowała rozmowę egzaminacyjną do interakcji komplementarnej⁴, czyli nierównorzędnej. Trzeba dodać, że czasami pojawia się chwilowa zmiana relacji na symetryczną, na przykład w przypadku wspólnego rozprawiania nad danym problemem przez egzaminatora i egzaminowanego (Majewska-Tworek 2014:29–40), jednakże nadrzędną pozycję w omawianej sytuacji komunikacyjnej należy przy-

² Gajda zauważa, że trudność sprawia odróżnienie gatunku od odmiany w obrębie danego gatunku.

³ Na podstawie analizy materiału rozmów egzaminacyjnych zebranego w ramach projektu GeWiss.

⁴ Podział interakcji na symetryczną i komplementarną wyróżniła Małgorzata Kita (1999).

pisać osobie egzaminującej (Majewska-Tworek 2014:49). Uczestnicy rozmowy egzaminacyjnej wchodzi w kooperację, jednak Majewska-Tworek (za Małgorzatą Kitą) zwraca uwagę na to, że „idealna kooperacja nie istnieje, a w określaniu tego typu relacji należy również uwzględnić cechę dominującą” (Majewska-Tworek 2014:40). Jednorodny typ kontaktu nie jest możliwy – jako przykład została podana rozmowa przyjacielska, podczas której może dojść do wzajemnego przerywania sobie przez rozmówców, co wiąże się z pewnym konfliktem (Majewska-Tworek 2014:40). Ilustracja nr 1 wskazuje uproszczony schemat rozmowy egzaminacyjnej.



Ilustracja 1: Uproszczony schemat rozmowy egzaminacyjnej

Źródło: opracowanie własne na podstawie rozdziału 2 („Rozmowa egzaminacyjna jako gatunek tekstu mówionego”) pracy Anny Majewskiej-Tworek „Niepłynność wypowiedzi w oficjalnej odmianie polszczyzny. Próba typologii” (2014).

2. Aspekt stylistyczny rozmowy egzaminacyjnej

Niesymetryczność uczestników rozmowy egzaminacyjnej prowadzi do pytania o to, czy występują znaczące różnice pomiędzy użyciem polszczyzny w rozmowie egzaminacyjnej przez egzaminatora i egzaminowanego. Jeśli tak, to jakie? Bożena Witosz stwierdza, że „każda interakcja [...] to spotkanie dwu indywidualności językowych” (Witosz 2006:17). Każdy uczestnik rozmowy egzaminacyjnej jest zatem indywidualnością językową (stosując określenie Witosz), która znajduje się w danej sytuacji komunikacyjnej i swój styl wypowiedzi dostosowuje (próbując lub powinna dostosować) w taki sposób, aby zrealizować dany gatunek. Porównanie elementów wypowiedzi osoby egzaminowanej i egzaminatora pozwoli zarysować obraz tego, jak dane osoby realizują odmianę oficjalną polszczyzny podczas rozmowy egzaminacyjnej.

W rozmowie egzaminacyjnej może przeplatać się kilka odmian funkcjonalnych polszczyzny. Ponieważ jest to sytuacja oficjalna, może pojawić się styl oficjalny w odmianie mówionej, jednakże nie jest wykluczone, że rozmówcy będą wplatać w rozmowę elementy charakterystyczne dla stylu potocznego. Ponadto mogą ujawnić się cechy charakterystyczne dla stylu naukowego – mimo że jest on zwykle klasyfikowany jako należący do odmiany pisanej polszczyzny (Wilkoń 2000) – w odmianie na przykład dydaktycznej (u egzaminatora) lub teoretycznej (zarówno u egzaminatora, jak i u egzaminowanego).

Aby zrealizować cel prezentowanych badań, przyjęto za podstawę porównania stylów egzaminowanego i egzaminatora kilka cech charakterystycznych odmian naukowej oraz oficjalnej polszczyzny (Wilkoń 2000):

- a) liczne struktury *proszę* + bezokolicznik;
- b) posługiwanie się strukturami nominalnymi (nominalność stylu);
- c) stosowanie rozbudowanych związków hipotaktycznych.

Cechy te powinny pozwolić uwydatnić różnice między odmianami funkcjonalnymi języka stosowanymi przez egzaminatora i egzaminowanego.

3. Wykorzystane metody i budowa korpusu

Do przeprowadzenia analizy zostały wykorzystane metody językoznawstwa korpusowego. Badane teksty podzielono na dwa subkorpusy: subkorpus E – wypowiedzi egzaminatorów oraz subkorpus S – wypowiedzi osób egzaminowanych, czyli studentów.

Subkorpusy zostały przygotowane na podstawie transkrypcji nagrań, wykonanych na potrzeby projektu GeWiss (2019) oraz opracowanych przez Annę Majewską-Tworek (2014) na potrzeby publikacji „Niepłynność wypowiedzi w oficjalnej odmianie polszczyzny. Próba typologii”. Analizie zostały poddane rozmowy egzaminacyjne 9 egzaminatorów z 32 studentami pochodzące z 25 nagrań. Rozmowy egzaminacyjne dotyczyły literatury, językoznawstwa, wiedzy o kulturze oraz glottodydaktyki. Tabela 1 zawiera zestawienie informacji o nagraniach.

Tabela 1: Zestawienie informacji o nagraniach

Numer nagrania	Osoby	Płeć	Temat	Czas trwania [hh:mm:ss]
001	E1 S1	mężczyzna kobieta	literatura	00:19:59
002	E1 S2	mężczyzna kobieta	literatura	00:15:33
003	E1 S3	mężczyzna kobieta	literatura	00:15:18

004	E1 S4	mężczyzna kobieta	literatura	00:13:41
005	E1 S5	mężczyzna kobieta	literatura	00:12:57
006	E1 S6	mężczyzna mężczyzna	literatura	00:20:29
007	E1 S7	mężczyzna kobieta	literatura	00:14:25
008	E1 S8	mężczyzna kobieta	literatura	00:21:28
009	E1 S9	mężczyzna kobieta	literatura	00:18:12
010	E1 S10	mężczyzna mężczyzna	literatura	00:22:03
012	E2 S11 S12 S13 S14	mężczyzna kobieta kobieta kobieta kobieta	literatura	00:35:00
013	E3 S15 S16 S17 S18 S19	mężczyzna kobieta kobieta kobieta kobieta mężczyzna	literatura	02:28:27
215	E4 E5 E6 S20	kobieta kobieta kobieta kobieta	glottodydaktyka	00:20:30
216	E4 E5 E6 S21	kobieta kobieta kobieta kobieta	glottodydaktyka	00:23:45
217	E7 S22	kobieta kobieta	literatura	00:17:28
218	E7 S23	kobieta kobieta	literatura	00:25:32
219	E7 S24	kobieta kobieta	literatura	00:12:43
220	E8 S25	kobieta mężczyzna	literatura	00:13:47
221	E9 S26	mężczyzna kobieta	wiedza o kulturze	00:09:36

222	E8 S27	kobieta kobieta	literatura	00:12:38
223	E5 S28	kobieta kobieta	językoznawstwo	00:22:03
224	E4 E5 E6 S29	kobieta kobieta kobieta kobieta	glottodydaktyka	00:23:00
225	E4 E5 E6 S30	kobieta kobieta kobieta kobieta	glottodydaktyka	00:17:30
226	E4 E5 E6 S31	kobieta kobieta kobieta kobieta	glottodydaktyka	00:20:35
227	E4 E5 E6 S32	kobieta kobieta kobieta kobieta	glottodydaktyka	00:22:37

Źródło: opracowanie na podstawie badań własnych.

Subkorpus E składa się z dokumentów zawierających wypowiedzi 9 egzaminatorów o łącznej objętości 21 679 tokenów. Subkorpus S zawiera dokumenty z wypowiedziami 32 studentów i liczy 41 296 tokenów. W tabeli nr 2 zostały zestawione dane dotyczące subkorpusów oraz osób, których teksty się na nie składają. Subkorpus S jest bardziej zrównoważony pod kątem udziału tekstów poszczególnych osób. Najwyższy udział tekstów jednej osoby wynosi 5,49%, a najniższy – 0,48%, ich średnia to 3,13%. W subkorpusie E najwyższy udział to 26,01%, a najniższy – 1,03%, ich średnia to 11,11%. Oznacza to, że wyniki otrzymywane dla wypowiedzi egzaminatorów mogą być bardziej narażone na wpływ cech idiolektalnych niektórych mówców.

Oba subkorpusy wyraźnie różnią się od siebie. Subkorpus wypowiedzi studentek ma niemal dwukrotnie większą objętość i zawiera wypowiedzi ponad trzykrotnie większej liczby osób. Wynika to ze sposobu zbierania materiału oraz z rzeczywistych różnic między populacjami studentów oraz egzaminatorów – niewielka liczba egzaminatorów egzaminuje zdecydowanie większą liczbę studentów. Różnice w objętości wypowiedzi poszczególnych mówców wynikają zaś przede wszystkim z tematyki egzaminu, jego formuły oraz poziomu wiedzy studenta. Asymetria między porównywanymi subkorpusami wynika zatem naturalnie z sytuacji pragmatycznej rozmowy egzaminacyjnej, jednakże objętość subkorpusów powinna pozwolić na otrzymanie miarodajnych wyników analiz ilościowych w wypadku większości badanych cech stylistycznych.

Tabela 2: Zestawienie udziału tekstów poszczególnych osób w subkorpusach

Subkorpus E			Subkorpus S		
Osoba	Liczba tokenów	Udział w subkorpusie [%]	Osoba	Liczba tokenów	Udział w subkorpusie [%]
E1	5638	26,01	S1	1785	4,33
E2	530	2,44	S2	883	2,14
E3	4056	18,71	S3	789	1,91
E4	1841	8,49	S4	1250	3,03
E5	4307	19,87	S5	1162	2,82
E6	1908	8,80	S6	1739	4,22
E7	2755	12,71	S7	1349	3,27
E8	224	1,03	S8	1019	2,47
E9	420	1,94	S9	1416	3,43
-			S10	2031	4,93
			S11	310	0,75
			S12	366	0,89
			S13	198	0,48
			S14	351	0,85
			S15	1160	2,81
			S16	2195	5,32
			S17	2229	5,41
			S18	2269	5,50
			S19	1481	3,44
			S20	1995	4,84
			S21	1015	2,46
			S22	1417	3,44
			S23	2265	5,49
			S24	1335	3,24
			S25	1748	4,24
			S26	612	1,48
			S27	1407	3,41
			S28	673	1,63
			S29	1482	3,59
			S30	937	2,27
			S31	885	2,15
			S32	1543	3,74
Suma	21 679	100,00	Suma	41 296	100,00

Źródło: opracowanie na podstawie badań własnych.

Na potrzeby analizy syntaktycznej i rozpoznania połączeń bezspójnikowych z wykorzystaniem pauz skorzystano z surowych nagrań i transkrypcji. Do badań wykorzystano oprogramowanie EXMARaLDA (2019). W większości pozostałych analiz wykorzystano narzędzie Korpusomat (2018). Korpusomat pozwala na lematyzację i tagowanie tekstów, co znacznie ułatwia analizę, oraz obsługuje składnię zapytań opartą na CQL oraz opracowaną na podstawie języka zapytań obsługiwanego przez wyszukiwarkę Poliqarp. Jej zaletą jest to, że korzysta ze szczegółowej i użytecznej w analizach klasyfikacji leksemów opracowanej na potrzeby korpusu IPI PAN (Woliński 2003). Aby umożliwić obróbkę tekstów przez Korpusomat oraz ich lematyzację, wyeksportowano transkrypty do postaci plików .txt oraz usunięto z nich oznaczenia pauz, zjawisk fonetycznych itp., pozostawiając jedynie wypowiedzany tekst tak, aby wypowiedzi mogły zostać poddane automatycznej analizie.

Oprócz EXMARaLDY oraz Korpusomatu skorzystano także z narzędzia Web-Sty (Piasecki/Walkowiak/Eder 2018) opracowanego przez konsorcjum CLARIN-PL do analizy stylometrycznej.

4. Struktury hipotaktyczne i parataktyczne

Style funkcjonalne polszczyzny charakteryzują się różnym wykorzystaniem zdań prostych i złożonych oraz zdań złożonych współrzędnie i podrzędnie. Ze względu na oralny charakter analizowanych tekstów oraz ich przynależność gatunkową trudno wyodrębnić takie jednostki organizacji tekstu jak zdania. Jednakże możliwa jest analiza łączliwości syntagmatycznej przy pomocy ustalenia frekwencji spójników, które mogą posłużyć za wskaźniki hipotaktyczności i parataktyczności oraz analizy połączeń bezspójnikowych.

a) Spójniki

Liczba spójników jest wywoływana w konkordancerze Korpusomatu z użyciem zapytania: [pos="(conj)"] dla parataksy i [pos="(comp)"] dla hipotaksy.

W subkorpusie E na 21 679 tokeny występuje: 1246 spójników współrzędnych reprezentujących 20 lematów oraz 593 spójniki podrzędne reprezentujące 18 lematów (tabela 3). Spójniki parataktyczne pojawiają się średnio co mniej więcej 17 tokenów i stanowią ich 5,75%, a hipotaktyczne – co mniej więcej 37 i stanowią 2,74% wszystkich tokenów w subkorpusie. Oznacza to, że spójniki hipotaktyczne są około 2,1 razy rzadsze od spójników parataktycznych.

W subkorpusie S na 41 296 tokenów pojawia się: 2077 spójników współrzędnych z 22 lematów oraz 1261 spójników podrzędnych z 29 lematów (tabela 4). Spójniki parataktyczne pojawiają się średnio co mniej więcej 20 tokeny

i stanowią około 5,03% wszystkich tokenów, a hipotaktyczne – co mniej więcej 33 i stanowią 3,05%. Spójniki hipotaktyczne są zatem około 1,65 razy rzadsze od parataktycznych.

W subkorpusie E jest bezwzględnie więcej spójników niż w subkorpusie S (E – 8,48%; S – 8,08%). Ponadto z danych wynika, że w subkorpusie E jest więcej spójników parataktycznych niż w subkorpusie S (E – 5,75%; S – 5,03%) przy odwrotnym wyniku w wypadku spójników hipotaktycznych (E – 2,74%; S – 3,05%). Jedyne różnica w stosunku spójników współrzędnych do podrzędnych wydaje się wyraźniejsza (E – 19:9, S – 15:9).

Otrzymane wyniki poddano trzem testom t-Studenta dla prób niezależnych – dla występowania spójników parataktycznych, hipotaktycznych oraz ich wzajemnego stosunku w wypowiedziach każdego mówcy. W wypadku 9 egzaminatorów i 32 studentów liczba stopni swobody wynosi 8, a wartość krytyczna rozkładu przy poziomie prawdopodobieństwa 0,05 wynosi 2,3060. Wynik testu dla spójników parataktycznych to 1,4028, dla spójników hipotaktycznych – 0,8033, dla ich stosunku – 0,9837. Rozkład żadnej z testowanych cech nie przekroczył wartości krytycznej, co oznacza, że obserwowane różnice nie są istotne statystycznie. A zatem porównywane subkorpusy nie różnią się między sobą w znaczny sposób pod kątem ilościowych różnic w występowaniu spójników.

W porównywanych subkorpusach wyraźniejsze od różnic ilościowych są różnice jakościowe. W wypowiedziach egzaminatorów najczęstszym spójnikiem współrzędnym jest *i* (33,95% wszystkich spójników parataktycznych), jednak również często występuje *a* oraz *ale* (23,84% i 19,02%). W wypowiedziach studentów przeważa *i*, ponieważ stanowi 55,42% wszystkich spójników współrzędnych, a żaden z pozostałych nie przekracza 10% (tabele 2 i 3 oraz wykresy 1 i 2). W wypadku spójników hipotaktycznych również widoczna jest dysproporcja. Najczęstszymi spójnikami są *że* (E – 33,22%; S – 47,42%) oraz *bo* (E – 28,67%; S – 8,8%), jednakże w subkorpusie S *że* jest wyraźnie częstsze, a *bo* jest jedynie nieznacznie częstsze od trzeciego w kolejności *więc* (tabela 4).

Tabela 3: Spójniki współrzędne i podrzędne w subkorpusie E

Subkorpus E					
Spójniki współrzędne [pos="(conj)"]			Spójniki podrzędne [pos="(comp)"]		
Spójnik	Liczba wystąpień	Udział [%]	Spójnik	Liczba wystąpień	Udział [%]
<i>i</i>	423	33,95	<i>że</i>	197	33,22
<i>a</i>	297	23,84	<i>bo</i>	170	28,67
<i>ale</i>	237	19,02	<i>więc</i>	47	7,93
<i>czyli</i>	81	6,50	<i>jeśli</i>	35	5,90
<i>jako</i>	58	4,65	<i>żeby</i>	29	4,89
<i>albo</i>	34	2,73	<i>jakby</i>	25	4,22

<i>czy</i>	30	2,41	<i>jeżeli</i>	19	3,20
<i>natomiast</i>	23	1,85	<i>gdyby</i>	14	2,36
<i>tylko</i>	20	1,61	<i>by</i>	10	1,69
<i>nie</i>	7	0,56	<i>ponieważ</i>	10	1,69
<i>inaczej</i>	6	0,48	<i>skoro</i>	8	1,35
<i>jednak</i>	6	0,48	<i>choć</i>	7	1,18
<i>niż</i>	5	0,40	<i>co</i>	6	1,01
<i>lub</i>	4	0,32	<i>zatem</i>	6	1,01
<i>ani</i>	3	0,24	<i>to</i>	4	0,67
<i>oraz</i>	3	0,24	<i>zanim</i>	3	0,51
<i>mianowicie</i>	3	0,24	<i>chociaż</i>	2	0,34
<i>zarówno</i>	3	0,24	<i>aczkolwiek</i>	1	0,17
<i>байд</i>	2	0,16	-	-	-
<i>ni</i>	1	0,08	-	-	-

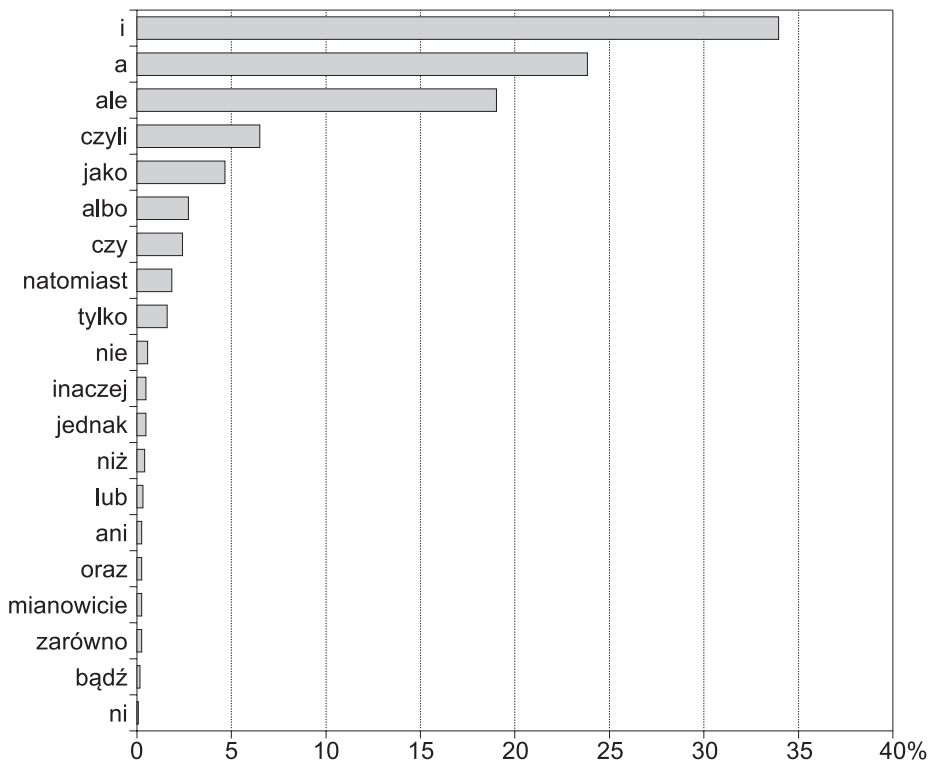
Źródło: opracowanie na podstawie badań własnych.

Tabela 4: Spójniki współrzędne i podrzędne w subkorpusie S

Subkorpus S					
Spójniki współrzędne [pos="(conj)"]			Spójniki podrzędne [pos="(comp)"]		
Spójnik	Liczba wystąpień	Udział [%]	Spójnik	Liczba wystąpień	Udział [%]
<i>i</i>	1151	55,42	<i>że</i>	598	47,42
<i>ale</i>	199	9,58	<i>bo</i>	111	8,80
<i>czyli</i>	181	8,71	<i>więc</i>	92	7,30
<i>a</i>	160	7,70	<i>jeżeli</i>	89	7,06
<i>jako</i>	85	4,09	<i>ponieważ</i>	88	6,98
<i>natomiast</i>	52	2,50	<i>żeby</i>	82	6,50
<i>jednak</i>	51	2,46	<i>jeśli</i>	63	5,00
<i>tylko</i>	46	2,21	<i>jakby</i>	48	3,81
<i>albo</i>	42	2,02	<i>chociaż</i>	25	1,98
<i>czy</i>	28	1,35	<i>gdyby</i>	11	0,87
<i>nie</i>	16	0,77	<i>co</i>	7	0,56
<i>niż</i>	14	0,67	<i>chociażby</i>	7	0,56
<i>oraz</i>	13	0,63	<i>aby</i>	6	0,48
<i>байд</i>	10	0,48	<i>to</i>	5	0,40
<i>jak</i>	9	0,43	<i>by</i>	4	0,32
<i>zarówno</i>	8	0,39	<i>aczkolwiek</i>	3	0,24
<i>lub</i>	4	0,19	<i>gdyż</i>	3	0,24

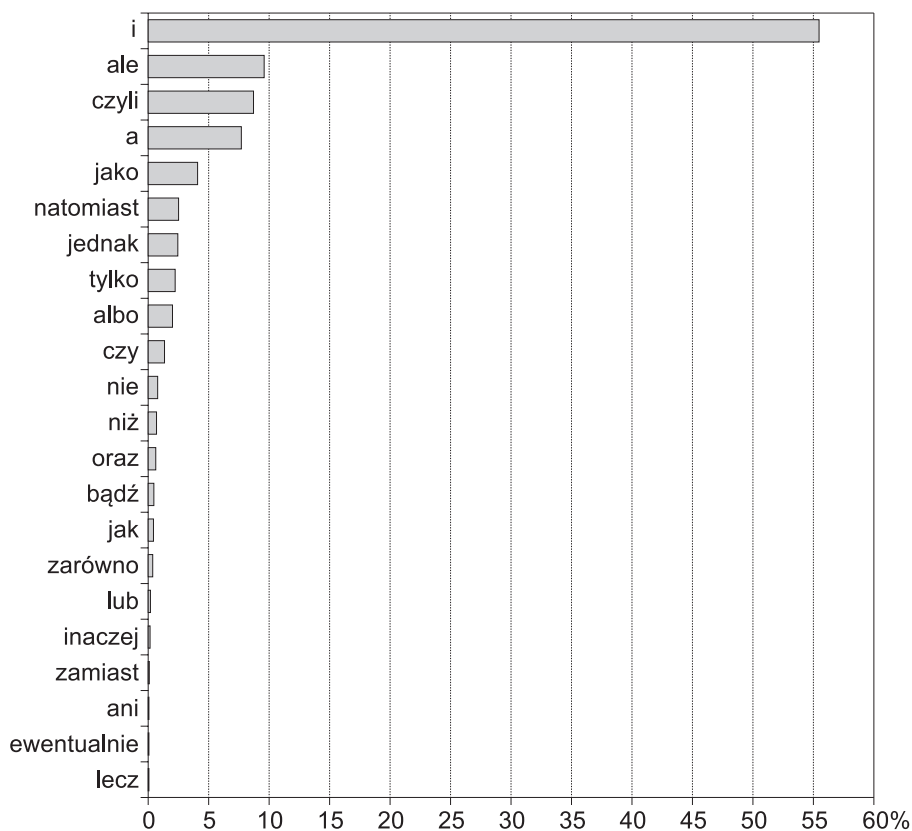
<i>inaczej</i>	3	0,14	<i>zatem</i>	3	0,24
<i>zamiast</i>	2	0,10	<i>choć</i>	2	0,16
<i>ani</i>	1	0,05	<i>bowiem</i>	2	0,16
<i>ewentualnie</i>	1	0,05	<i>im</i>	2	0,16
<i>lecz</i>	1	0,05	<i>jako</i>	2	0,16
-	-	-	<i>skoro</i>	2	0,16
-	-	-	<i>choćby</i>	1	0,08
-	-	-	<i>dopóki</i>	1	0,08
-	-	-	<i>kiedy</i>	1	0,08
-	-	-	<i>iż</i>	1	0,08
-	-	-	<i>jeżeli</i>	1	0,08
-	-	-	<i>póki</i>	1	0,08

Źródło: opracowanie na podstawie badań własnych.



Wykres 1: Procentowy udział poszczególnych spójników wśród spójników parataktycznych w subkorpusie E

Źródło: opracowanie na podstawie badań własnych.



Wykres 2. Procentowy udział poszczególnych spójników wśród spójników parataktycznych w subkorpusie S

Źródło: opracowanie na podstawie badań własnych.

b) Pauzy w funkcji syntaktycznej

Oprócz spójników w tekście mówionym mogą występować sygnały łączliwości syntaktycznej w postaci pauz segmentujących wypowiedź i występujących między syntagmami. Na podstawie analizy audytywnej⁵ zidentyfikowano w obu subkorpusach pauzy⁶ mogące pełnić funkcję syntaktyczną, będące sygnałem łączliwości bezspójnikowej. Są to pauzy, które mogłyby być zastąpione spójnikiem. W identyfikowaniu takich pauz ważne jest pomijanie tych, które jedynie towarzyszą spójnikom.

⁵ Analiza – Kamil Wabnic.

⁶ Omawiane pauzy w przykładach są zaznaczone **boldem**.

Takie pauzy syntaktyczne mogą być puste, na przykład: *w literaturze była cenzura represyjna (.) po wydaniu książki cenzor mógł ingerować (0.5) w innych dziedzinach jak kino (0.3) albo radio przewencyjna* oraz *ja bym powiedziała że był realistyczny (0.7) często właśnie dotyczył takich codziennych potocznych spraw.* Część pauz jest również wypełnionych: *spory wpływ na to miało opublikowanie h° listy katyńskiej listy (0.2) osób które zostały zamordowane w katyniu (0.7) °h (3.0) yyy po wojnie również (.) yyy (0.8) ((cmoka)) yyy ten motyw yyy pojawia się w poezji.*

Pauz syntaktycznych w subkorpusie E jest 339, a w subkorpusie S – 1236. W stosunku do liczby tokenów jest to kolejno: 1,56% i 2,99%. Jednocześnie w subkorpusie E stosunek liczby pauz syntaktycznych do spójników wynosi około 1 do 5, czyli 7 do 40, a w subkorpusie S – około 3 do 8, czyli 15 do 40.

Stosunek liczby pauz do liczby tokenów w obu subkorpusach poddano testowi t-Studenta o takich samych parametrach jak w wypadku analizy spójników. Wynik testu to 4,5213, czyli przekracza on wartość krytyczną – 2,3060, co oznacza, że jest on istotny statystycznie. Zatem różnica we frekwencji pauz syntaktycznych w obu subkorpusach jest cechą, która je od siebie odróżnia.

c) Podsumowanie analizy syntaktycznej

W subkorpusie E jest bezwzględnie więcej spójników niż w subkorpusie S (E – 8,48%; S – 8,08%), jednakże ta różnica jest nieznaczna i w podobnym stopniu jest widoczna we frekwencji spójników parataktycznych i hipotaktycznych. Większe rozbieżności między subkorpusami występują w wypadku analizy frekwencji pauz syntaktycznych, których jest wyraźnie więcej w wypowiedziach studentów.

Można również zaryzykować wniosek, że liczba pauz równoważy nieznaczny niedobór formalnych wyznaczników spójności syntaktycznej w wypowiedziach osób egzaminowanych (tabela 5), gdyż sprawia, że sumaryczna frekwencja pauz i spójników w subkorpusie S jest wyższa niż w subkorpusie E. Ogólna wyższa liczba wskaźników połączeń syntaktycznych między wypowiedziami może wynikać z tego, że średnio wypowiedzi studentów są dłuższe niż wypowiedzi egzaminatorów – co jest konsekwencją sytuacji pragmatycznej rozmowy egzaminacyjnej – a zatem występuje więcej połączeń wewnątrz nich.

Ponadto pauzy można uznać za sygnały spójności parataktycznej, ponieważ nie hierarchizują one spajanych wypowiedzeń, nie zawierają również informacji o ich stosunku logicznym. Można zatem uznać wypowiedzi studentów za mniej spójne niż wypowiedzi egzaminatorów, co może wynikać z przygotowywania wypowiedzi w czasie rzeczywistym oraz stresu egzaminacyjnego i niesymetrycznej relacji rozmówców.

Tabela 5: Połączenia syntaktyczne między wypowiedziami w subkorpusach E i S

Połączenia syntaktyczne między wypowiedziami				
	Subkorpus E		Subkorpus S	
	Liczba	Stosunek do tokenów bez spójników	Liczba	Stosunek do tokenów bez spójników
Spójniki hipotaktyczne	593	około 30:1000	1261	około 33:1000
Spójniki parataktyczne	1246	około 63:1000	2077	około 55:1000
Pauzy syntaktyczne	339	około 17:1000	1236	około 33:1000
Suma spójników i pauz syntaktycznych	2178	około 110:1000	4574	około 121:1000
Liczba tokenów bez spójników	19840	1000:1000	37958	1000:1000

Źródło: opracowanie na podstawie badań własnych.

5. Nominalność i abstrakcyjność stylu

Zarówno styl oficjalny, jak i naukowy charakteryzują się stosowaniem licznych konstrukcji nominalnych, których przejawem jest duża liczba rzeczowników w stosunku do czasowników. Ponadto jedną z cech stylu naukowego jest jego abstrakcyjność, można ją przeciwstawić konkretności, która charakteryzuje styl potoczny. W rozmowie egzaminacyjnej ze względu na jej szczególny charakter ścierają się zatem tendencje do konkretyzacji poprzez odniesienie do sytuacji pragmatycznej oraz do abstrahowania ze względu na naukową tematykę. Za wyznacznik abstrakcyjności lub konkretności tekstu może posłużyć analiza zastępowalności czasowników przez gerundia. Styl silnie czasownikowy i pozbawiony gerundiów jest konkretny, a rzeczownikowy – abstrakcyjny.

Liczba rzeczowników jest wywoływana w konkordancerze Korpusomatu poprzez zapytanie [pos="(subst|depr|ger)"], a czasowników – [(pos="(fin|bedzie|praet|impt|imps|inf|winien|pred)"&!orth="(to)"]].

W subkorpusie E na 21 679 tokenów przypada 6310 rzeczowników (29,11% tokenów), w tym 164 gerundia (0,76%) oraz 3292 czasowniki (15,19%). Na jeden czasownik przypada zatem około 1,91 rzeczownika.

W subkorpusie S na 41296 tokenów występuje 11 345 rzeczowników (27,47%), w tym 416 gerundiów (1,01%) oraz 6040 czasowników (14,63%). Oznacza to, że na jeden czasownik przypada około 1,88 rzeczownika.

Dane liczbowe dla obu subkorpusów są bardzo podobne, jedynie różnica w liczbie rzeczowników jest wyraźniejsza, ponieważ czasowniki w subkorpusie E są liczniejsze o 1,64 punktu procentowego. Jednakże dwa testy t-Studenta o stałych parametrach dla tych subkorpusów przeprowadzone dla frekwencji rzeczowników i czasowników wykazały, że obserwowane różnice ilościowe nie są istotne statystycznie. Wynik testu dla rzeczowników to 0,9709, a dla czasowników – 0,8421, czyli nie przekraczają one wartości krytycznej wynoszącej 2,3060.

Wyraźniejsze różnice niż w danych ilościowych pojawiają się w warstwie semantycznej czasowników. Tabela 6 zawiera zestawienie piętnastu najliczniej występujących czasowników w obu subkorpusach. W obu grupach tekstów najczęściej jest realizacji lematu *być* (E – 22,75%; S – 28,08%) oraz *mieć* (E – 5,04%; S – 4,97%), co jest naturalne, gdyż słowa te pełnią funkcje gramatyczne. W wypowiedziach studentów słowa te wyraźnie przeważają, a żaden z pozostałych czasowników nie stanowi więcej niż 4% (trzęcie *móc* – 3,08%).

W wypowiedziach egzaminatorów trzecim najliczniejszym czasownikiem jest *prosić* – stanowi on 4,89% wszystkich czasowników. Czasownik ten w subkorpusie E występuje niemal wyłącznie⁷ w pierwszej osobie liczby pojedynczej czasu teraźniejszego *proszę* i pełni wyraźne funkcje pragmatyczne. Występuje on w funkcji dyrektywnej w połączeniu z innymi czasownikami, na przykład: *proszę wziąć, proszę dać, proszę odczytać, proszę dodać, proszę powiedzieć, proszę usiąść* itp. Czasownik ten występuje również w funkcji adresatywnej *proszę pani* lub samodzielnie jako zachęta dla studenta do udzielenia odpowiedzi. W subkorpusie S czasownik *prosić* występuje jedynie 8 razy (0,13% wszystkich czasowników), z czego 5 w cytatach pytań egzaminacyjnych. Wynika to z faktu, że studenci nie wydają poleceń egzaminatorom, a jeśli pojawiają się jakiegokolwiek dyrektywne akty mowy, to mają one bardzo niską moc illokucyjną i występują na przykład w postaci niebezpośrednich pytań. Świadczy o tym również niemal całkowity brak wystąpień bezpośrednich zwrotów do egzaminatora – lematy *pan* i *pani* w takiej funkcji w subkorpusie S występują tylko 9 razy (na 40 wszystkich wystąpień, przy 576 wystąpieniach w subkorpusie E).

W obu subkorpusach wśród najliczniejszych czasowników znalazły się takie, które występują w wypowiedzeniach o charakterze metatekstowym, oraz dotyczące procesów myślowych i stosunku nadawcy do wypowiedzi, na przykład: *chodzić* – *chodzi mi o to; rozumiem, że chodzi o to; powiedzieć* – *można powiedzieć, że; ja bym powiedziała, że; zacząć* – *chciałabym zacząć od tego, wiedzieć* – *nie wiem*.

⁷ *Proszę* – 159 razy, *prosimy* – 1, *prosiła* – 1.

Tabela 6: Najliczniejsze czasowniki w subkorpusach E i S

Najliczniejsze czasowniki [(pos="(fin bedzie praet impt imps inf winien pred)"&!orth="(to"))]					
Subkorpus E			Subkorpus S		
Lemat	Liczba	Procent czasowników	Lemat	Liczba	Procent czasowników
<i>być</i>	749	22,75	<i>być</i>	1696	28,08
<i>mieć</i>	166	5,04	<i>mieć</i>	300	4,97
<i>prosić</i>	161	4,89	<i>móc</i>	186	3,08
<i>powiedzieć</i>	117	3,55	<i>znaczyć</i>	174	2,88
<i>móc</i>	106	3,22	<i>chodzić</i>	119	1,97
<i>mówić</i>	76	2,31	<i>wiedzieć</i>	117	1,94
<i>chodzić</i>	65	1,97	<i>powiedzieć</i>	103	1,71
<i>musieć</i>	59	1,79	<i>można</i>	94	1,56
<i>można</i>	55	1,67	<i>musieć</i>	77	1,27
<i>dziękować</i>	52	1,58	<i>mówić</i>	62	1,03
<i>wiedzieć</i>	41	1,25	<i>wydawać</i>	62	1,03
<i>znaczyć</i>	36	1,09	<i>chcieć</i>	55	0,91
<i>myśleć</i>	35	1,06	<i>pamiętać</i>	55	0,91
<i>pamiętać</i>	35	1,06	<i>myśleć</i>	51	0,84
<i>trzeba</i>	25	0,76	<i>pisać</i>	46	0,76

Źródło: opracowanie na podstawie badań własnych.

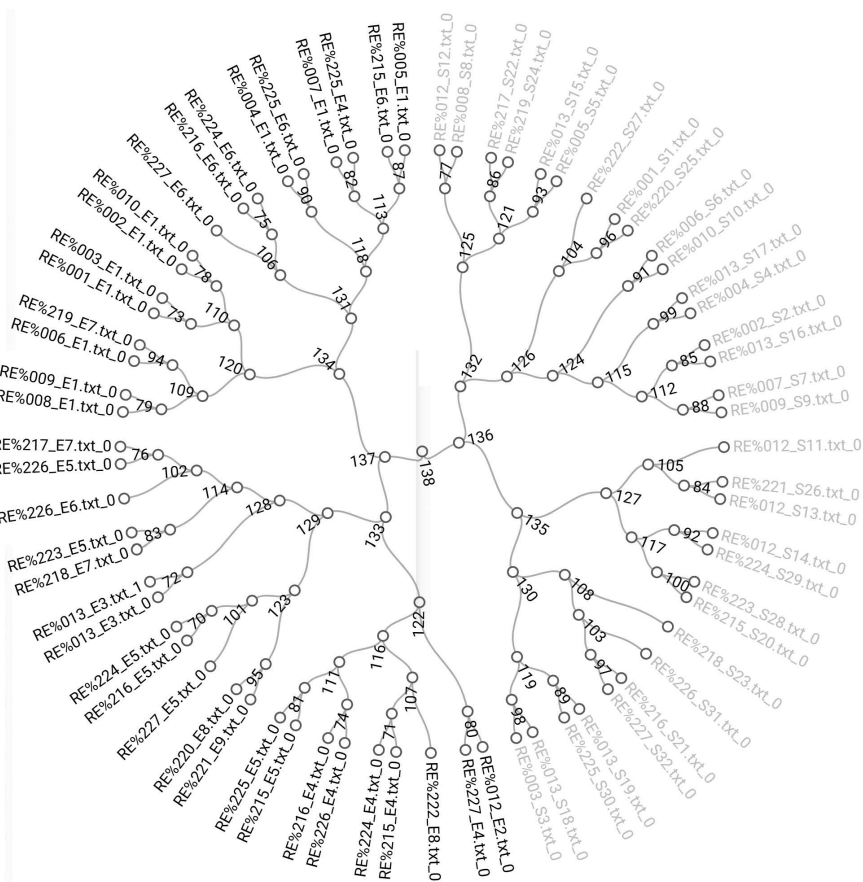
6. Analiza stylometryczna – 100 najczęstszych słów

Metody stylometryczne tradycyjnie wykorzystywane są do analizy autorstwa tekstów, jednakże wyniki badań eksperymentalnych pokazują, że stylometria może być z powodzeniem wykorzystywana w analizie podobieństwa tekstów pod kątem reprezentowanych gatunku oraz stylu (Rybicki 2014).

Do analizy wypowiedzi uczestników rozmów egzaminacyjnych wykorzystano system analizy podobieństwa tekstów WebSty (Piasecki/Walkowiak/Eder 2018). Wypowiedzi egzaminatorów oraz studentów podzielono na dokumenty według nagrań (zob. tabela 1) – z każdego transkryptu wyeksportowano po jednym pliku tekstowym dla każdego mówcy. W ten sposób otrzymano 32 dokumenty z wypowiedziami studentów oraz 37 dokumentów z wypowiedziami egzaminatorów.

Jedną z podstawowych metod analizy stylometrycznej jest porównanie w analizowanych tekstach najczęściej występujących słów. Przy analizie tego typu istotny jest wybór odpowiedniej liczby porównywanych słów – badacze przyjmują zasadę, że im liczniejsza jest lista, tym większe znaczenie w analizie

ma tematyka tekstu (Cheesman/Flanagan/Rybicki/Stephan 2016). Na potrzeby porównania podobieństwa wypowiedzi studentów i egzaminatorów wybrano jedynie sto najczęstszych lematów. W wyniku analizy algorytm wyodrębnił dwie grupy podobieństwa tekstów, które pokrywają się z podziałem na wypowiedzi studentów i egzaminatorów (ilustracja 2), a zatem najczęściej występujące lematy w porównywanych grupach tekstów są istotnie różne.



Ilustracja 2. Diagram podobieństwa tekstów

Źródło: aplikacja WebSty (Piasecki/Walkowiak/Eder 2018).

7. Wnioski

Przeprowadzone analizy wykazały liczne podobieństwa stylistyczne między badanymi subkorpusami. Styl tekstów w żadnym z subkorpusów nie jest silnie nominalny – liczba rzeczowników w stosunku do czasowników nie przekracza 2:1.

Zastępowalność czasowników gerundiami także jest znikoma. Brak istotności zaobserwowanych różnic potwierdziły również testy statystyczne. Różnice w warstwie semantycznej czasowników również są znikome – wyraźnie dominuje *być* oraz *mieć*, a także czasowniki występujące w wypowiedzeniach metatekstowych oraz opisujących stany mentalne rozmówców. Wyraźną różnicą między subkorpusami jest obecność struktur z *proszę* – często w funkcji dyrektywnej – w wypowiedziach egzaminatorów, podczas gdy w wypowiedziach studentów nie występują one niemalże wcale.

Różnice między korpusami występują też w zakresie dystrybucji środków odpowiadających za spójność syntaktyczną. Żaden z subkorpusów nie zawiera licznych rozbudowanych związków hipotaktycznych, przeważa parataksa, która jednak jest realizowana w różny sposób. W wypowiedziach studentów większe znaczenie ma łączliwość bezspójnikowa, która występuje między wypowiedziami i uobecnia się w postaci pauz. Również spójniki używane przez studentów charakteryzują się mniejszą różnorodnością – wyraźnie dominuje *i* oraz *że*. Może to wynikać z mniejszego uporządkowania wypowiedzi oraz z przygotowywania jej na bieżąco, a wyższa frekwencja pauz – z większej średniej długości wypowiedzi. Co ważne – istotności różnic we frekwencji spójników nie potwierdziły testy statystyczne, w przeciwieństwie do różnic we frekwencji pauz – te okazały się statystycznie istotne.

Również na istotne różnice między subkorpusami wskazuje wynik analizy stylometrycznej przeprowadzonej narzędziem WebSty. Algorytm bezbłędnie podzielił wypowiedzi na dwie grupy na podstawie stu najczęściej występujących lematów.

Analiza obu subkorpusów wykazała jedynie nieznaczne różnice ilościowe między badanymi zjawiskami, jednakże pozwoliła ujawnić różnice jakościowe, które mogą stać się przedmiotem dalszych, bardziej szczegółowych analiz. Zważywszy, że zestawienie występowania czasowników wykazało wiele dyrektywnych użyć słowa *prosić* w wypowiedziach egzaminatorów, szczególnie owocna może okazać się na przykład analiza pragmatyczna, która obejmowałaby porównanie ilościowe i jakościowe aktów mowy występujących w badanych grupach tekstów.

Bibliografia

- CHEESMAN Tom / FLANAGAN Kevin / RYBICKI Jan / STEPHAN Thiel, 2016, Six maps of translations of Shakespeare, w: Wiggin B./Macleod C. (red.), Un/translatable. New Maps for Germanic Literatures, Evanston.
- EXMARaLDA, 2019, <https://exmaralda.org/en/>.
- GAJDA Stanisław, 1982, Podstawy badań stylistycznych nad językiem naukowym, Wrocław.
- GeWiss, 2019, <https://gewiss.uni-leipzig.de/index.php?id=home>.

- Grammis. Grammatisches Informationssystem: Propädeutische Grammatik, Pausen: <https://grammis.ids-mannheim.de/progr/mm/4821>.
- KIERAŚ Witold / KOBYLŃSKI Łukasz / OGRODNICZUK Maciej, 2018, Korpusomat — a tool for creating searchable morphosyntactically tagged corpora, in: Computational Methods in Science and Technology 24(1), s. 21–27, <http://dx.doi.org/10.12921/cmst.2018.0000005>.
- KITA Małgorzata, 1999, Porozmawiajmy o rozmowie. Kryteria typologii rozmowy jako interakcji werbalnej, w: Stylistyka 8, s. 119–131.
- MAJEWSKA-TWOREK Anna, 2014, Niepłynność wypowiedzi w oficjalnej odmianie polszczyzny. Próba typologii, Wrocław.
- PIASECKI Maciej / WALKOWIAK Tomasz / EDER Maciej, 2018, Open Stylometric System WebSty: Towards multilingual and multipurpose workbench, w: Piasecki M. (red.), Selected Papers from the CLARIN Annual Conference 2017, Budapest.
- RYBICKI Jan, 2014, Pierwszy rzut oka na stylometryczną mapę literatury polskiej, w: Teksty Drugie 2, s. 106–128.
- TARAS Bożena, 2003, Słowo do słowa, zrobi się rozmowa, czyli kilka spostrzeżeń na temat rozmowy, w: Kita M./Grzenia J. (red.), Porozmawiajmy o rozmowie. Lingwistyczne aspekty dialogu, Katowice.
- WILKOŃ Aleksander, 2000, Typologia odmian językowych współczesnej polszczyzny, Katowice.
- WILKOŃ Aleksander, 2003, Gatunki mówione, w: Kita M./Grzenia J. (red.), Porozmawiajmy o rozmowie. Lingwistyczne aspekty dialogu, Katowice.
- WITOSZ Bożena, 2006, O badaniach stylów konwersacyjnych z perspektywy współczesnej stylistyki, w: Witosz B. (red.), Style konwersacyjne, Katowice.
- WOLIŃSKI Marcin, 2003, System znaczników morfosyntaktycznych w korpusie IPI PAN, w: Polonica 12–13, s. 39–55, <http://nlp.ipipan.waw.pl/CORPUS/znakowanie.pdf>.
- ŻYDEK-BEDNARCZUK Urszula, 1994, Struktura tekstu rozmowy potocznej, Katowice.

Oral exam – selected features of style of utterances of examiners and students. Corpus analysis

The purpose of the article is to compare selected features of the style of utterances of professors and students in an oral exam as a communication situation. The research material consists of 25 recordings of oral exam (9 examiners with 32 students). They come from a corpus collected as part of GeWiss – a study project on the spoken scientific language. The texts were divided into two subcorpora: E (examiners) and S (students). Corpus linguistics methods were used in analysis. Several characteristic features of scientific and official styles were compared: numerous structures *proszę* + infinitive; nominal structures (nominal style); extensive hypotaxis. The analysis showed numerous stylistic similarities between the examined subcorpora. The style of none of the texts in the subcorpora is strongly nominal. A clear difference between the subcorpora is the presence of structures with the word *proszę* – it appears in the utterances of examiners, while in the utterances of students it is almost non-existent. The distribution of means responsible for cohesion in both subcorpora is different (parataxis is more common than hypotaxis but is implemented differently); also, there are differences in lists of one hundred most frequently used lexemes in the subcorpora – these differences allow us to distinguish these texts with tools for automatic style similarity analysis.

Keywords: oral exam, style, pause, corpus linguistics, examiner, examined person, stylometry.