

Łukasz Rogoziński
Grzegorz Zarzeczny
Uniwersytet Wrocławski

Mówca AD_1489 w świetle wybranych parametrów ilościowych

W artykule przedstawiamy przykładowe analizy, na jakie pozwala zbudowany na Uniwersytecie Wrocławskim podkorpus teksów polszczyzny mówionej w domenie akademickiej (Rogoziński, Zarzeczny, Rogozińska 2012) stworzony w ramach projektu GeWiss. W pierwszej części tekstu skrótowo omawiamy założenia i kształt korpusu oraz charakteryzujemy analizowany materiał. Druga część jest poświęcona wybranym wskaźnikom dotyczącym języka mówionego. Część trzecia to omówienie danych empirycznych.

1. Podkorpus polszczyzny akademickiej (GeWiss POL)

W ramach projektu GeWiss (GesprocheneWissenschaftssprache Kontrastiv) powstał niemiecko-angielsko-polski kontrastywny korpus obejmujący łącznie 120 godzin języka mówionego w odmianie akademickiej. Dwa gatunki obecne w korpusie to egzaminy ustne (PG) oraz referaty naukowe: eksperckie (EV – od stopnia magistra) i studenckie (SV). Tematycznie nagrania obejmują subdyscypliny tradycyjnej filologii, tj. literaturoznawstwo, językoznawstwo oraz dydaktykę (Fandrych i in. 2009; Fandrych, Meißner, Slavcheva 2012). W skład podkorpusu polonistycznego (GeWiss POL) wchodzi łącznie dwadzieścia godzin nagrań w języku polskim, w tym dziesięć godzin egzaminów, pięć godzin referatów eksperckich oraz pięć godzin referatów studenckich (Rogoziński, Zarzeczny, Rogozińska 2012). Wszystkie nagrania zostały przetranskrybowane według zasad transkrypcji minimalnej (Minimaltranskript) systemu GAT 2 (Selting i in. 2009) z pewnymi zmianami wynikającymi z dostosowania do języka polskiego oraz wyszukiwania korpusowego (Rogoziński, Zarzeczny 2012). Całość prac, tj. transkrypcja wraz z anotacją oraz metadanymi, została wykonana z użyciem programu EXMARaLDA (Schmidt, Wörner 2009).

Korpus GeWiss stanowił już podstawę kilku wstępnych analiz – głównie o charakterze jakościowym – obejmujących zjawiska specyficzne dla polszczyzny mówionej: zjawiska fonetyczne (Tworek 2011), nie płynność (Majewska-Tworek 2011), autokorekty (Koman 2011) czy wypełniacze (Rogoziński, Zarzeczny 2011). Tutaj chcielibyśmy omówić kilka parametrów ilościowych przydatnych we wstępnym opisie zgromadzonych danych.

Na potrzeby przykładowej analizy wybraliśmy jednego z mówców, którzy pojawiają się w wielu nagraniach. Jako że wszystkie dane wrażliwe osób uczestniczących w projekcie są anonimizowane, mówca ten jest oznaczony w korpusie kodem AD_1489. Bierze on udział w siedmiu zarejestrowanych zdarzeniach komunikacyjnych, w tym sześć z nich (PG_PL_215, PG_PL_216, PG_PL_224, PG_PL_225, PG_PL_226, PG_PL_227¹) to jednorodne w kształcie rozmowy egzaminacyjne z zakresu dydaktyki języka polskiego jako obcego (II rok studiów magisterskich). Stałą rolą komunikacyjną mówcy AD_1489 oraz dwojga innych uczestników (AB_1490 i UD_1491) jest rola egzaminatora, przy czym pierwszy z nich pełni funkcję lidera – zawsze rozpoczyna zdarzenie komunikacyjne, zadając pierwsze pytanie bądź udzielając głosu innemu egzaminatorowi, a także czuwa nad przebiegiem całości egzaminu, stymulując jego przebieg. Tabela 1 zawiera przykładowe frazy AD_1489 sterujące przebiegiem całego zdarzenia komunikacyjnego.

Tabela 1. Wybrane komunikaty mówcy AD_1489 sterujące zdarzeniem komunikacyjnym

TREŚĆ	NAGRANIE
Sekwencje otwierające – samodzielne	
zaczniemy egzamin i pierwsze (0,4) pytanie (1,3) proste (0,8) pomoce	PG_PL_215
dobrze to wobec tego (0,7) yyy moje pytanie dla pani dotyczy	PG_PL_216
zaczniemy egzamin i mam pierwsze pytanie do pani takie (1,0)	PG_PL_226
Sekwencje otwierające – przekazanie głosu	
pierwsze pyta pierwsze pytanie proszę pani (.) doktor	PG_PL_224
dobrze ja dziękuję i bardzo proszę	PG_PL_224
°h dobrze dobrze dziękuję i jeszcze teraz sprawy prak (.) tyk	PG_PL_224
Sekwencje sterujące	
jedno się zgadza z tego co pani powiedziała	PG_PL_216
a czemu pani pod stół im kazała wchodzić ((śmiech))	PG_PL_215
to znaczy o co chodzi tutaj jak pani mówi zachód to o czym pani myśli	PG_PL_224

¹ Pierwsze dwie litery informują o gatunku: w tym przypadku PG – skrót niemieckiego Prüfungsgespräch (rozmowa egzaminacyjna / egzamin ustny). Kolejne dwie to oznaczenie kraju, w którym została dokonana rejestracja nagrania: w tym przypadku PL – Polska. Cyfry na końcu nazw nagrań to numery porządkowe dla poszczególnych rozmów egzaminacyjnych.

Pozostali mówcy (AMS_1488, MK_1486, AW_1483, AM_1584, BS_1485, RP_1487) biorą udział w poszczególnych zdarzeniach komunikacyjnych w roli egzaminowanych. Wskazane transkrypty stały się podstawą wybranych analiz ilościowych, przy czym dla mówcy AD_1489 przyjęto dwa punkty odniesienia: zarówno dane dotyczące egzaminatorów, jak i egzaminowanych.

2. Charakterystyka poszczególnych nagrań

Tabela 2 prezentuje przyporządkowanie studentów do poszczególnych nagrań, czas wykorzystany przez mówców na wypowiedzi (podany w minutach i sekundach), proporcjonalne zaangażowanie mówców w całe zdarzenie komunikacyjne (wyrażone w procentach), łączną długość wypowiedzi wszystkich mówców biorących udział w danym nagraniu oraz właściwą długość samego nagrania.

Tabela 2. Zestawienie długości (czasu) wypowiedzi mówców w nagraniach

Nagranie	PG_PL_215		PG_PL_216		PG_PL_224		PG_PL_225		PG_PL_226		PG_PL_227	
	czas	%	czas	%	czas	%	czas	%	czas	%	czas	%
AD_1489	3:29	16	4:33	17	3:39	17	2:48	16	3:58	19	6:55	27
AB_1490	1:42	8	2:58	11	1:04	5	2:05	12	2:51	13	3:30	13
UD_1491	2:16	10	4:27	17	3:40	17	1:49	10	3:32	17	0:34	2
AMS_1488	14:59	67										
MK_1486			14:13	54								
AW_1483					13:37	62						
AM_1584							11:19	63				
BS_1485									10:54	51		
RP_1487											15:07	58
Suma	22:26		26:11		22:00		18:01		21:15		26:06	
Czas nagrania	20:30		23:45		23:00		17:30		20:35		22:37	

Wartości czasowe przypisane poszczególnym mówcom zostały uzyskane podczas ręcznej analizy każdej z linii werbalnych składających się na partyturę anotacyjną całego nagrania. Linie werbalne składają się z segmentów, które w układzie horyzontalnym obrazują przebieg zaangażowania mówcy w określonym momencie zdarzenia komunikacyjnego, w układzie wertykalnym zaś – stosunek zaangażowania wobec innych mówców. Taki system anotacyjny pozwala na wyodrębnianie tych segmentów, w których mówcy są współaktywni werbalnie lub parawerbalnie (*overlap*). Stąd też zsumowa-

ny czas wypowiedzi wszystkich uczestników nagrania jest dłuższy od czasu trwania samego nagrania. Różnica tych dwu wartości to suma długości wszystkich momentów zdarzenia komunikacyjnego, w których głos zabierało więcej osób.

3. Wybrane parametry analityczne

Specyfika korpusów tekstów mówionych na tle standardowych korpusów tekstów pisanych ujawnia się przede wszystkim w anotacji różnych zjawisk niewerbalnych współtworzących tekst mówiony w interakcji komunikacyjnej. Stopień dokładności powyższych anotacji zależy, oczywiście, od przyjętych kryteriów. W korpusie GeWiss POL uznano, że anotowane w linii werbalnej jako osobne jednostki są między innymi pauzy, wypełniacze (*fillers*) czy inne samoistne (tj. niebędące elementami wypowiedzianej frazy) zjawiska parawerbalne, np. sygnały aprobaty lub dezaprobaty, śmiech itp. (por. Rogoziński, Zarzeczny 2012). Wszystkie one są traktowane jako pełnoprawne jednostki komunikacyjne, jako że pełnią istotną funkcję w interakcji, nie są zaś wyłącznie „zaburzeniami” mowy (por. np. Linell 2005; O’Connell, Kowal 2008). W tabeli 3 znajduje się zestawienie liczbowe wszystkich jednostek komunikacyjnych (JK) z podziałem na jednostki parawerbalne (JP) oraz wyrazy² dla analizowanego mówcy. Ostatnia kolumna – stopień werbalizacji³ wypowiedzi (SW) – wyraża procentowy stosunek wyrazów do wszystkich jednostek komunikacyjnych.

Tabela 3. Zestawienie jednostek komunikacyjnych mówcy AD_1489

AD_1489	JK	JP	P	SW (w %)
PG_PL_215	515	86	429	83
PG_PL_216	712	41	671	94
PG_PL_224	645	60	585	91
PG_PL_225	440	54	386	88
PG_PL_226	584	91	493	84
PG_PL_227	1045	111	934	89
Razem	3941	443	3498	89

² Wyraz z konieczności rozumiany tu jest jako wyraz graficzny, tj. ciąg znaków pomiędzy spacjami. Analogicznie pod względem wyodrębnienia zostały zdefiniowane jednostki parawerbalne.

³ Termin *w e r b a l i z a c j a* definiujemy tu jako ‘budowanie komunikatu wyłącznie za pomocą słów’, bez udziału elementów niewerbalnych itp.

Zestawiając powyższe dane z czasem trwania wypowiedzi w poszczególnych sytuacjach komunikacyjnych, można wyznaczyć dwa dodatkowe parametry charakteryzujące mówcę. Jest to uśrednione tempo komunikowania (TK) oraz uśrednione tempo werbalizacji (TW) liczone odpowiednio w jednostkach komunikacji oraz wyrazach realizowanych w ciągu jednej minuty.

Tabela 4. Tempo wypowiedzi i tempo werbalizacji mówcy AD_1489

AD_1489	Czas (w s)	JK	W	TK/min	TW/min
PG_PL_215	209	515	429	148	123
PG_PL_216	273	712	671	156	147
PG_PL_224	219	645	585	177	160
PG_PL_225	168	440	386	157	138
PG_PL_226	238	584	493	147	124
PG_PL_227	415	1045	934	151	135
Razem	1522	3941	3498	155	138

Oczywiście, parametry te są wyłącznie orientacyjne – przede wszystkim ze względu na różną długość poszczególnych wyrazów. Docelowo, po anotacji morfologicznej zawartości korpusu, wskaźnik ten należałoby uszczegółowić, jako podstawę obliczeń przyjmując nie wyraz, lecz sylabę. Niemniej jednak już teraz, ze względu na jednorodność gatunkową i tematyczną analizowanych zdarzeń komunikacyjnych, a także – pewną porównywalność wszystkich mówców pod względem wykształcenia, podane parametry mogą zostać wykorzystane do ogólnej charakterystyki mówcy. Analogiczne dane dla pozostałych mówców zawiera tabela 7 (w Załączniku).

Kolejne parametry są związane z niewyrazowymi realizacjami wokalicznymi rejestrowanymi w korpusie. Pierwszy z nich to wypełniacze (Y), czyli niewerbalne wokalizacje. Charakteryzują się one dużą wariantywnością, jeśli chodzi o rzeczywiście artykułowane dźwięki, choć u poszczególnych mówców miewają bardziej ujednolicone brzmienia. Na potrzeby tak dużego korpusu w transkryptach oznaczane są niezależnie od rzeczywistej realizacji jako [yyy]. Drugi to analogiczne wokalizacje (M) – oznaczane w transkryptach jako [mmm] – różniące się od poprzednich tym, że są nośnikami jakiejś konkretnej treści, np. przytaknięcia, zaprzeczenia, wątpliwości. Można je więc nazwać sygnałami recepcyjnymi. Prototypowe realizacje [yyy] różnią się od analogicznych [mmm] (niezależnie od właściwości artykulatoryjnych) właściwościami akustycznymi. Pierwsze mają niezmienny kontur intonacyjny, tj. brak zmian w wysokości po wyabstrahowaniu częstotliwo-

ści podstawowej (tonu krtaniowego) F0. Drugie mają wyraźnie zarysowany kontur intonacyjny – po uzyskaniu i odtworzeniu F0 nie ma zazwyczaj problemu z odtworzeniem znaczenia niesionego przez dany sygnał recepcyjny. Dla charakterystyki mówcy przyjmujemy uśrednioną liczbę obu zjawisk na minutę.

Ostatnim wskaźnikiem, który może posłużyć do charakterystyki mówcy, jest czas milczenia (CM), czyli udział w wypowiedzi elementów komunikatu bez webralizacji i sygnałów recepcyjnych (zob. tab. 5). Jego podstawę stanowi łączny czas trwania pauz (wartość podana jest w sekundach) oraz mikropauz. Te ostatnie – w transkrypcie oznaczane zgodnie z przyjętymi konwencjami transkrypcyjnymi jako (.) – są to pauzy, których długość nie przekraczała 0,2 sekundy, a tu została uśredniona do 0,1 sekundy. Należy tu zaznaczyć, że do powyższych zjawisk nie są wliczane wszystkie fragmenty zdarzenia komunikacyjnego, w których mówca nie zabierał głosu, ale jedynie integralne części jego poszczególnych tur komunikacyjnych.

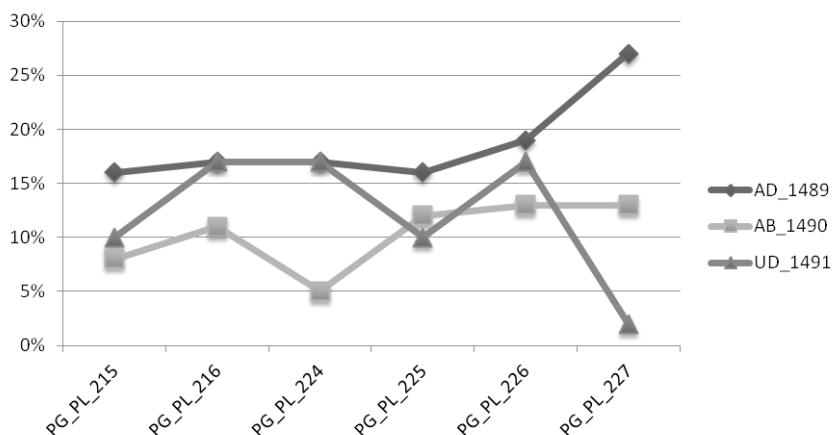
Tabela 5. Niewyrazowe realizacje wokaliczne oraz czas milczenia mówcy AD_1489

AD_1489	Czas (w s)	Y	Y/min	M	M/min	Pauzy i (.)		CM (w %)
PG_PL_215	209	15	4,3	11	3,2	18,5	1,1	9
PG_PL_216	273	15	3,3	2	0,4	10,5	0,8	4
PG_PL_224	219	16	4,4	6	1,6	3,6	0,9	2
PG_PL_225	168	9	3,2	4	1,4	15,3	0,3	9
PG_PL_226	238	17	4,3	6	1,5	22,5	1,2	10
PG_PL_227	415	47	6,8	12	1,7	16,6	1,4	4
Razem	1522	119	4,7	41	1,6	87,0	5,7	6

Zbiorcze dane na temat powyższych parametrów dla pozostałych uczestników poszczególnych zdarzeń komunikacyjnych zawiera tabela 8 (w Załączniku).

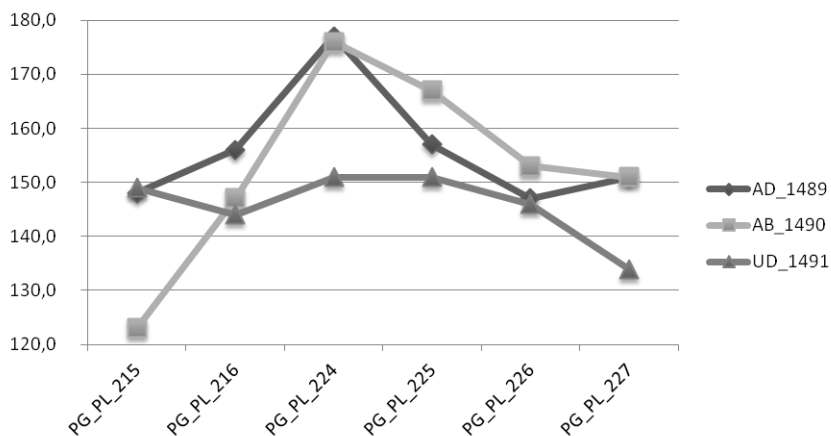
4. Mówca AD_1489 w świetle parametrów ilościowych

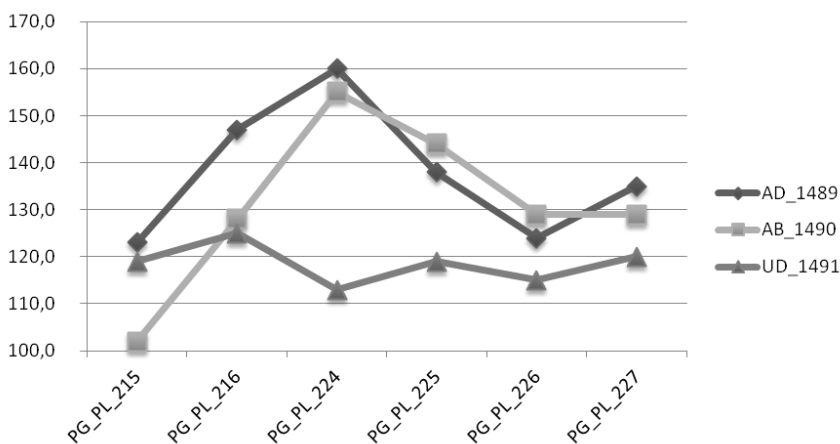
We wszystkich analizowanych rozmowach egzaminacyjnych najwięcej czasu (od połowy do 2/3 nagrania) zajmują, rzecz jasna, wypowiedzi egzaminowanych studentów (zob. tab. 2). Spośród egzaminatorów rolę lidera, tj. najaktywniejszego z mówców, odgrywa właśnie AD_1489. W każdym zdarzeniu komunikacyjnym procentowy udział tego mówcy jest najwyższy – w dwu *ex aequo* z mówcą UD_1491 (zob. wykres 1).



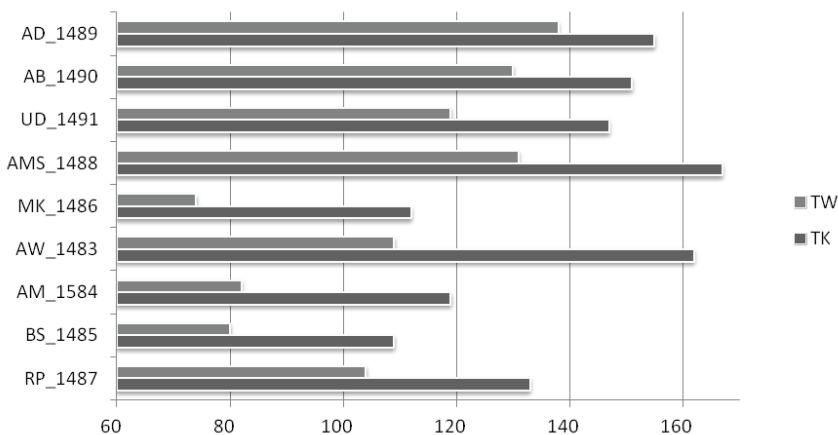
Wykres 1. Procentowy udział egzaminatorów w poszczególnych nagraniach

W swoich wypowiedziach mówca AD_1489 uzyskał (po uśrednieniu wszystkich egzaminów) najszybsze tempo komunikacji oraz tempo werbalizacji spośród wszystkich egzaminatorów, a także – najszybsze tempo werbalizacji spośród wszystkich mówców analizowanych nagrań. Zestawienie tych dwu parametrów (wykresy 2 i 3) wskazuje, że mówca AD_1489 realizował w czasie jednej minuty najwięcej wyrazów, a jednocześnie stosunkowo niewiele elementów parawerbalnych. Niewielką różnicę między tymi parametrami do pewnego stopnia można utożsamiać z p ł y n n o s c i ą wypowiedzi.



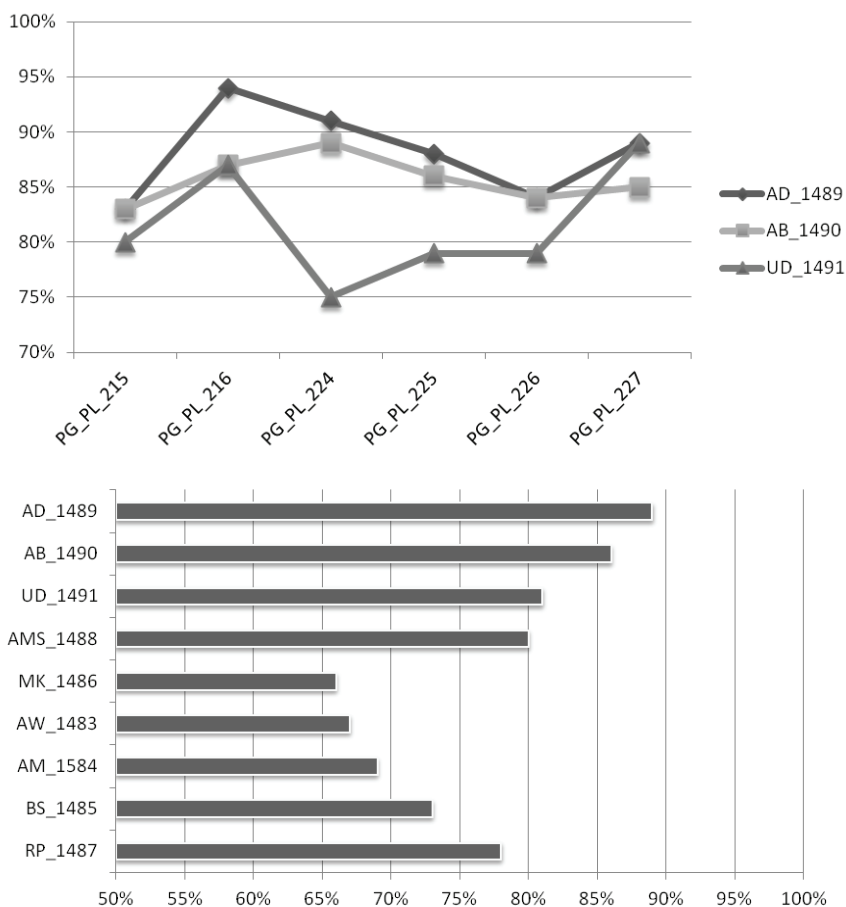


Wykres 2. Parametry TK i TW egzaminatorów w poszczególnych nagraniach



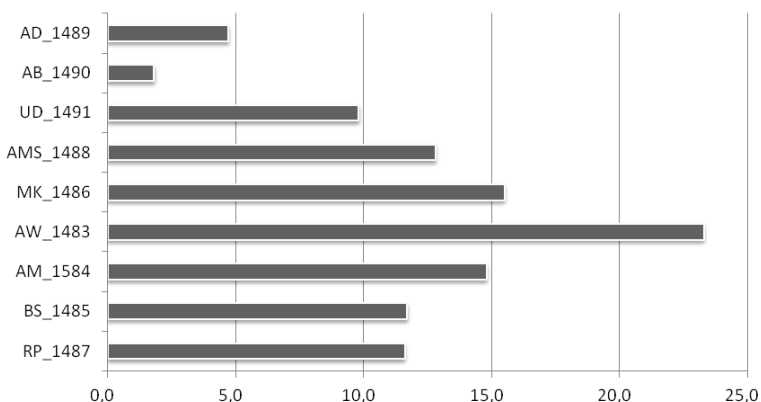
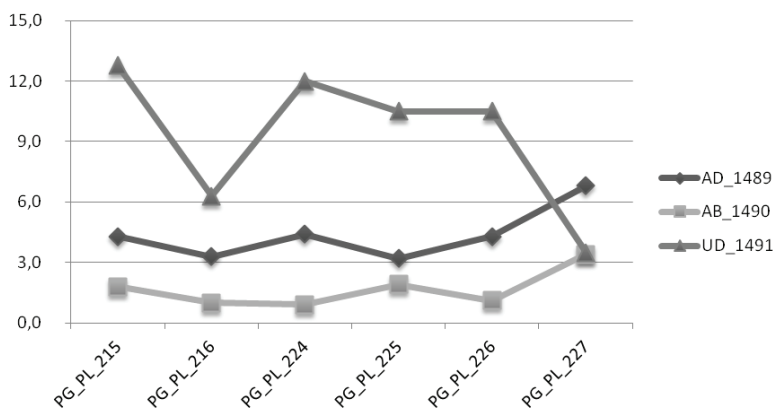
Wykres 3. Uśrednione parametry TK i TW poszczególnych mówców

Znajduje to także odzwierciedlenie w uzyskanym przez analizowanego mówcę najwyższym stopniu werbalizacji wypowiedzi (wykres 4). Parametr ten dla egzaminatorów waha się w granicach 80–90%, studenci zaś w swych wypowiedziach używają od 20 do nawet 34% elementów parawerbalnych. Dla AD_1489 jest to zaledwie 11%.



Wykres 4. Parametr SW (w poszczególnych nagraniach oraz uśredniony)

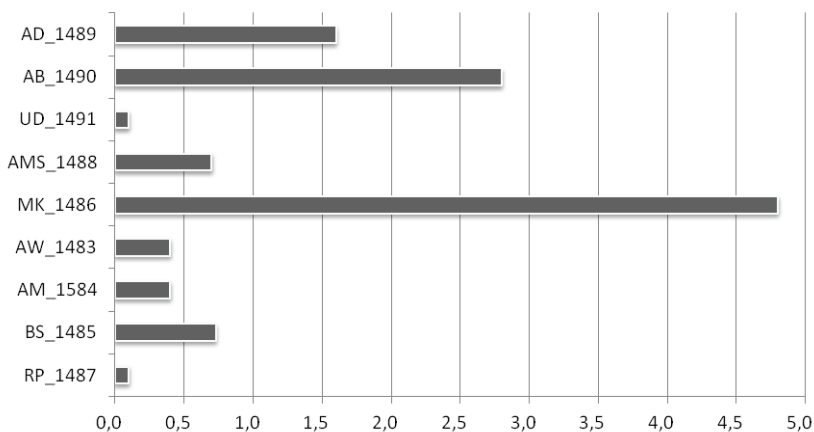
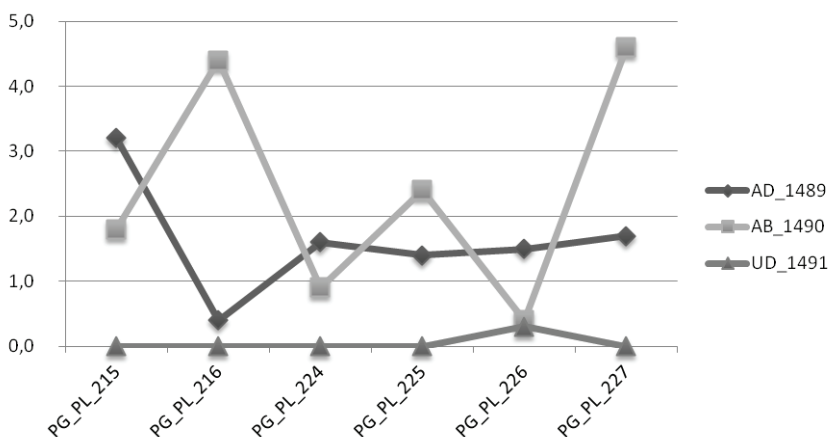
O płynności wypowiedzi mówcy AD_1489 świadczy również stosunkowo rzadkie (średnio 4,7/min) występowanie wypełniaczy (Y). Najmniej tego typu elementów (1,8/min) stosuje inny z egzaminatorów – AB_1490. Zupełnie inaczej wartość tego parametru wygląda dla wypowiedzi studentów – zawsze powyżej 10/min – co, rzecz jasna, wynika z innej roli komunikacyjnej (zob. wykres 5). Wśród najistotniejszych czynników warto wymienić niższy stopień przygotowania oraz pewności wypowiedzi, a także stosowanie wypełniaczy jako elementu utrzymywania się przy głosie (*floor holding*) (Rogoziński, Zarzeczny 2011).



Wykres 5. Parametr Y (w poszczególnych nagraniach oraz uśredniony)

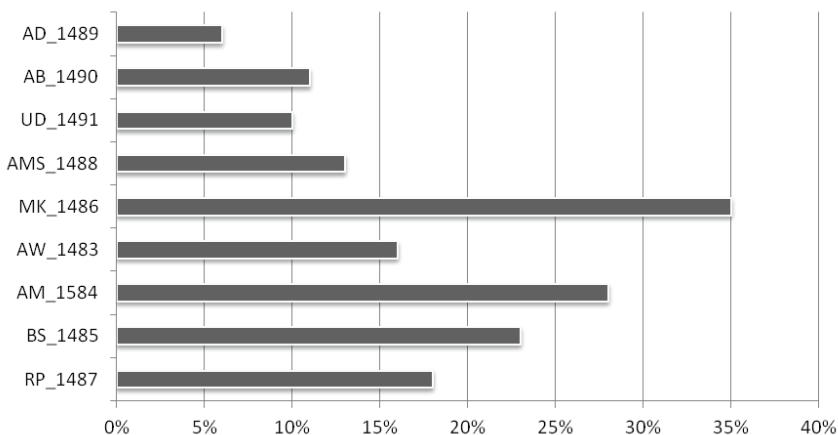
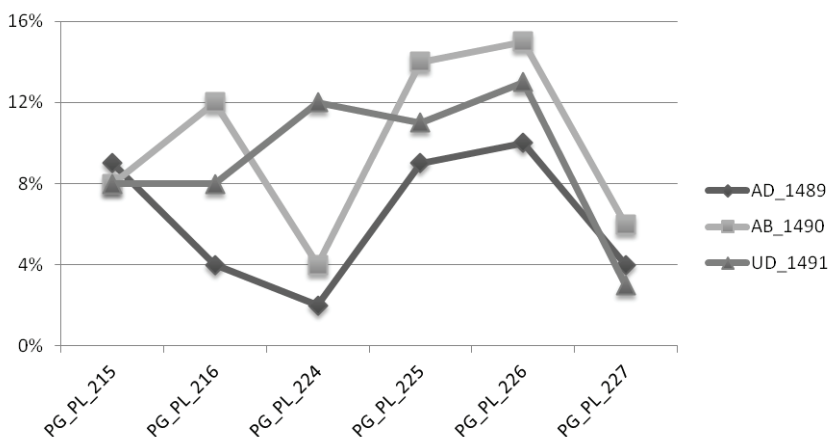
Mówca AD_1489 używa średnio 1,6/min sygnałów recepcyjnych (M). Większą interakcyjnością charakteryzuje się mówca AB_1490, natomiast trzeci egzaminator (UD_1491) używa sygnałów recepcyjnych zdecydowanie najmniej. Wynika to z roli w zdarzeniu komunikacyjnym – część egzaminu przeprowadzana przez tego mówcę nie ma charakteru interakcyjnego w takim stopniu jak jego pozostałe elementy⁴. Wśród studentów ze zwykle niewielkiego parametru M wyłamuje się mówca MK_1486. Znajduje to uzasadnienie w zestawieniu z wartościami parametrów TK oraz CM – w obu przypadkach największymi pośród wszystkich mówców (zob. wykres 6).

⁴ W części tej referowane, analizowane i oceniane są praktyki zawodowe egzaminowanych.



Wykres 6. Parametr M (w poszczególnych nagraniach oraz uśredniony)

Czas milczenia (wykres 7) poszczególnych egzaminatorów jest stosunkowo stabilny (średnio 6–11% wypowiedzi), przy czym dla mówcy AD_1489 przeciętnie jest on najkrótszy lub – w pierwszym i ostatnim nagraniu – zbliżony z wynikami pozostałych mówców. Zdecydowanie inaczej natomiast wygląda w przypadku poszczególnych egzaminowanych: tutaj można wskazać co najmniej trzy grupy: najbliższą egzaminatorom (10–20% milczenia; AMS_1488, AW_1483), przeciętną (20–30%; RP_1487, BS_1485, AM_1584) oraz najdalszą (>30%; MK_1486).



Wykres 7. Parametr CM (w poszczególnych nagraniach oraz uśredniony)

Tę wyraźną różnicę można by spróbować uogólnić już nie tylko na analizowany materiał, ale po prostu na zachowania komunikacyjne poszczególnych mówców. Taki test pozwoliłby też do pewnego stopnia porównać, w jakim stopniu wcześniejsze, odnoszące się wyłącznie do analizowanych nagrań dane, byłyby statystycznie istotne dla poszczególnych mówców w ogóle. W związku z tym przetestowaliśmy hipotezę H_0 o równości średnich długości pauz poszczególnych mówców. Jako że wszystkie próby z badanych populacji mają licznosci powyżej 30, a rozkłady są nieznane, sprawdzianem testu była statystyka t :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Średnie i odchylenia standardowe wyliczone z danych zawiera tabela 9 (w Załączniku). Poziom ufności ustaliliśmy na 95%. Dane dotyczące poszczególnych mówców porównaliśmy parami. Istotności różnic statystycznych zawiera tabela 6 (E - różnica nieistotna statystycznie, R - różnica istotna statystycznie).

Tabela 6. Istotności różnic statystycznych pauz poszczególnych mówców

	AD	AB	UD	AMS	MK	AW	AM	BS	RP
AD_1489	X	E	E	E	R	E	R	R	R
AB_1490		X	R	E	R	E	R	R	R
UD_1491			X	E	R	E	R	R	R
AMS_1488				X	R	E	R	R	R
MK_1486					X	R	R	R	R
AW_1483						X	R	R	R
AM_1584							X	E	E
BS_1485								X	E
RP_1487									X

Na przyjętym poziomie ufności średnie długości pauz stosowanych przez mówcę MK_1486 zdecydowanie różnią się od pauz pozostałych mówców. Spośród pozostałych egzaminowanych paury RP_1487, BS_1485 oraz AM_1584 są istotnie różne od pauz AMS_1488 oraz AW_1483. W przypadku pozostałych relacji – również względem egzaminatorów – hipotezy o równości średnich odrzucić nie sposób. Można zatem przyjąć, że uszeregowanie oparte na wcześniejszej analizie danych daje się (w podanym zakresie) rozciągnąć na ogólne zachowania komunikacyjne mówców.

5. Podsumowanie

Wskazane parametry ilościowe zdarzeń komunikacyjnych wydają się stosunkowo przydatnym narzędziem analitycznym. Przede wszystkim pozwalają do pewnego stopnia zobiektywizować niektóre intuicyjne kategorie

opisu (np. p ł y n n o ś ć w y p o w i e d z i). Umożliwiają też doprecyzowanie wybranych ustaleń pochodzących z oglądu jakościowego danych (np. rola lidera zdarzenia komunikacyjnego). Dodatkowo test wybranego parametru pokazuje, że podane wyniki statystyk deskrypcyjnych dają również szansę indukcji na ogólniejszy poziom opisu.

Ograniczenia zastosowanych metod wydają się równie oczywiste. Przede wszystkim przy dużym nakładzie pracy pozwalają wskazać jedynie mniej istotne, w niektórych przypadkach nawet banalne, elementy charakterystyki zdarzenia komunikacyjnego i jego uczestników. Równie ważnym ograniczeniem jest, wskazany już, deskrypcyjny charakter – dopiero statystyczne uogólnienie pozwala na więcej niż przeprowadzony tu opis interakcji w konkretnych zdarzeniach komunikacyjnych. Wreszcie przykład znaczącej różnicy parametru *M* u jednego z egzaminatorów pokazuje, że bez wzięcia pod uwagę danych jakościowych, interpretacja wyników ilościowych jest niedokładna lub wręcz fałszywa.

Na podstawie przytoczonych wniosków można stwierdzić, że w badaniach języka mówionego sensowne wydaje się stosowanie ilościowej parametryzacji zdarzenia komunikacyjnego w kombinacji z oglądem jakościowym, przy czym punktem wyjścia byłby jednak opis ilościowy. Jego interpretację warto następnie zweryfikować (i ewentualnie skorygować), uwzględniając dane jakościowe. Taki opis zdarzenia komunikacyjnego można traktować jako dogodny punkt wyjścia dalszych badań korpusowych, np. analizy leksykalnej czy bardziej rozbudowanych – stylistycznej, gatunkowej lub dyskursywnej.

Bibliografia

- Fandrych Christian i in., 2009, *Gesprochene Wissenschaftssprache kontrastiv: Deutsch im Vergleich zum Englischen und Polnischen. Vorstellung eines gemeinsamen Forschungsvorhabens*, „*Studia Linguistica*” 28, s. 7–30.
- Fandrych Christian, Meißner Cordula, Slavcheva Adriana, 2012, *The GeWiss Corpus: Comparing Spoken Academic German, English and Polish*, [w:] *Multilingual corpora and multilingual corpus analysis*, red. Thomas Schmidt, Kai Wörner, John Benjamins, Amsterdam, s. 319–337.
- Koman M., 2011, *Selbstkorrekturen bei Deutsch sprechenden Polen. Eine Fallstudie zur Gesprochenen Sprache im Germanistikstudium*, [w:] *Germanistische Linguistik im interdisziplinären Gefüge I*, red. Iwona Bartoszewicz, Joanna Szczęk, Artur Tworek, ATUT/Neisse Verlag, Wrocław–Dresden, s. 125–131.
- Linnell Per, 2005, *The written language bias in linguistics: Its nature, origins and transformations*, Routledge, London.

- Majewska-Tworek Anna, 2011, *Disfluenz als ein universelles Merkmal des spontanen Redens*, [w:] *Germanistische Linguistik im interdisziplinären Gefüge I*, red. Iwona Bartoszewicz, Joanna Szczęk, Artur Tworek, ATUT/Neisse Verlag, Wrocław–Dresden, s. 115–124.
- O’Connell Daniel C., Kowal Sabine, 2008, *Communicating with One Another. Toward a Psychology of Spontaneous Spoken Discourse*, Springer, New York.
- Rogoziniński Łukasz, Zarzeczny Grzegorz, 2011, *Verwendung des Phonetischen Korpus in der Diskursanalyse. Signaluntersuchung und Forschungshypothese*, [w:] *Germanistische Linguistik im interdisziplinären Gefüge I*, red. Iwona Bartoszewicz, Joanna Szczęk, Artur Tworek, ATUT/Neisse Verlag, Wrocław–Dresden, s. 105–113.
- Rogoziniński Łukasz, Zarzeczny Grzegorz, 2012, *Adapting GAT 2 conventions for the transcription of spoken Polish as part of the GeWiss project*, „*Studia Linguistica*” 31 [w druku].
- Rogoziniński Łukasz, Zarzeczny Grzegorz, Rogozińska Marta, 2012, *Aufbereitung und Erstellung eines Fachkorpus der gesprochenen Sprache am Beispiel des polnischen Teils des GeWiss-Korpus, Motoren der heutigen (germanistischen) Linguistik*, red. Edyta Błachut, Adam Gołębiowski, ATUT/Neisse Verlag, Wrocław–Dresden, s. 107–120.
- Schmidt Thomas, Wörner Kai, 2009, *EXMARaLDA – Creating, analyzing and sharing spoken language corpora for pragmatic research*, „*Pragmatics*” 19, s. 565–582.
- Selting Margret i in., 2009, *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)*, „*Ge-sprächsforschung – Online-Zeitschrift zur verbalen Interaktion*” 10, s. 353–402.
- Tworek Artur, 2011, *Zur Methodologie der phonetischen Analyse gesprochener Texte aus der korpuslinguistischen Sicht*, [w:] *Germanistische Linguistik im interdisziplinären Gefüge I*, red. Iwona Bartoszewicz, Joanna Szczęk, Artur Tworek, ATUT/Neisse Verlag, Wrocław–Dresden, s. 97–103.

Speaker AD_1489 in the light of selected quantitative parameters

Summary

In this paper we present exemplary analysis which are feasible by using (sub)corpus of Polish spoken academic discourse that was created at the University of Wrocław as part of the GeWiss project (Rogozinski, Zarzeczny, Rogozińska 2012). In the first part of the article we briefly present the assumptions and the structure of corpus as well as we describe the substance that is core of this paper. The second part is dedicated to selected indicators which relate to spoken language. The third part is description and interpretation of the empirical data.

Keywords: GeWiss, spoken academic discourse, spoken language corpus, quantitative analysis.

Załącznik

Tabela 7. Jednostki komunikacyjne oraz stopień i tempo werbalizacji u pozostałych mówców

Mówca	Nagranie	Czas (w s)	JK	JP	W	SW (w %)	TK/ min	TW/ min
AB_1490	PG_PL_215	102	209	35	174	83	123	102
	PG_PL_216	178	437	56	381	87	147	128
	PG_PL_224	64	186	21	165	89	174	155
	PG_PL_225	125	348	49	299	86	167	144
	PG_PL_226	171	435	68	367	84	153	129
	PG_PL_227	210	529	78	451	85	151	129
	Razem	850	2144	307	1837	86	151	130
UD_1491	PG_PL_215	136	337	67	270	80	149	119
	PG_PL_216	267	639	81	558	87	144	125
	PG_PL_224	220	553	137	416	75	151	113
	PG_PL_225	109	274	57	217	79	151	119
	PG_PL_226	212	517	110	407	79	146	115
	PG_PL_227	34	76	8	68	89	134	120
	Razem	978	2396	460	1936	81	147	119
AMS_1488	PG_PL_215	899	2470	502	1968	80	165	131
MK_1486	PG_PL_216	853	1593	547	1046	66	112	74
AW_1483	PG_PL_224	817	2202	723	1479	67	162	109
AM_1584	PG_PL_225	679	1348	416	932	69	119	82
BS_1485	PG_PL_226	654	1189	318	871	73	109	80
RP_1487	PG_PL_227	907	2011	446	1565	78	133	104

Tabela 8. Parametry parawerbalne u pozostałych mówców

Mówca	Nagranie	Czas (w s)	Y	Y/min	M	M/ min	Pauzy i (.) (w s)	CM (w %)
AB_1490	PG_PL_215	102	3	1,8	3	1,8	8,4	8
	PG_PL_216	178	3	1,0	13	4,4	21,7	12
	PG_PL_224	64	1	0,9	1	0,9	2,5	4
	PG_PL_225	125	4	1,9	5	2,4	17	14
	PG_PL_226	171	3	1,1	1	0,4	26,2	15
	PG_PL_227	210	12	3,4	16	4,6	13,6	6
	Razem	850	26	1,8	39	2,8	89,4	11

UD_1491	PG_PL_215	136	29	12,8	0	0,0	10,8	8
	PG_PL_216	267	28	6,3	0	0,0	21,1	8
	PG_PL_224	220	44	12,0	0	0,0	26,8	12
	PG_PL_225	109	19	10,5	0	0,0	11,7	11
	PG_PL_226	212	37	10,5	1	0,3	27,1	13
	PG_PL_227	34	2	3,5	0	0,0	0,9	3
	Razem	978	159	9,8	1	0,1	98,4	10
AMS_1488	PG_PL_215	899	192	12,8	10	0,7	120,1	13
MK_1486	PG_PL_216	853	221	15,5	68	4,8	297,1	35
AW_1483	PG_PL_224	817	317	23,3	5	0,4	134,7	16
AM_1584	PG_PL_225	679	168	14,8	4	0,4	191	28
BS_1485	PG_PL_226	654	128	11,7	8	0,7	152,1	23
RP_1487	PG_PL_227	907	175	11,6	1	0,1	165,9	18

Tabela 9. Dane statystyczne pauz poszczególnych mówców

	Liczność	Średnia długość pauzy (w s)	Odchylenie standardowe
AD_1489	155	0,56	0,3103353
AB_1490	143	0,59	0,4465997
UD_1491	183	0,50	0,3216245
AMS_1488	224	0,50	0,4946388
MK_1486	203	1,40	1,8566448
AW_1483	246	0,50	0,5882534
AM_1584	199	0,94	1,2826381
BS_1485	138	1,08	1,5722019
RP_1487	177	0,92	0,7394252