

Jakościowa i ilościowa analiza słownictwa oraz jednostek quasi- i nieleksykalnych wypowiedzi w dialogach zadaniowych „origami”

Qualitative and quantitative analysis of lexical, quasi- and non-lexical components of utterances in task-oriented “origami” dialogues

Michał Szczyszek^{a, b}, Maciej Karpiński^{a, c}

^aINTERDYSCYPLINARNE CENTRUM PRZETWARZANIA MOWY I JEZYKA UAM*

^bINSTYTUT FILOLOGII POLSKIEJ UAM

UL. A. FREDRY 10, 61-701 POZNAŃ

^cINSTYTUT JEZYKOZNAWSTWA UAM

*AL. NIEPODLEGŁOŚCI 4, 61-874 POZNAŃ

szczysze@amu.edu.pl, maciej.karpinski@amu.edu.pl

Abstract

In this text, an analysis of the lexical material obtained from the recordings of Polish task-oriented dialogues is presented. Some findings regarding quasi- and non-lexical units are also mentioned. The dialogue task involved the reconstruction of a figure made of paper. It was carried out in two settings: Mutual visibility and limited visibility. The analyses of the vocabulary are focused on finding its specific features and the differences between the structures of the vocabularies of instruction givers and instruction followers in the two settings. Frequency lists as well as most frequent bigrams extracted from the material are discussed in terms of their relatedness to the settings, the roles of the speakers and some other factors.

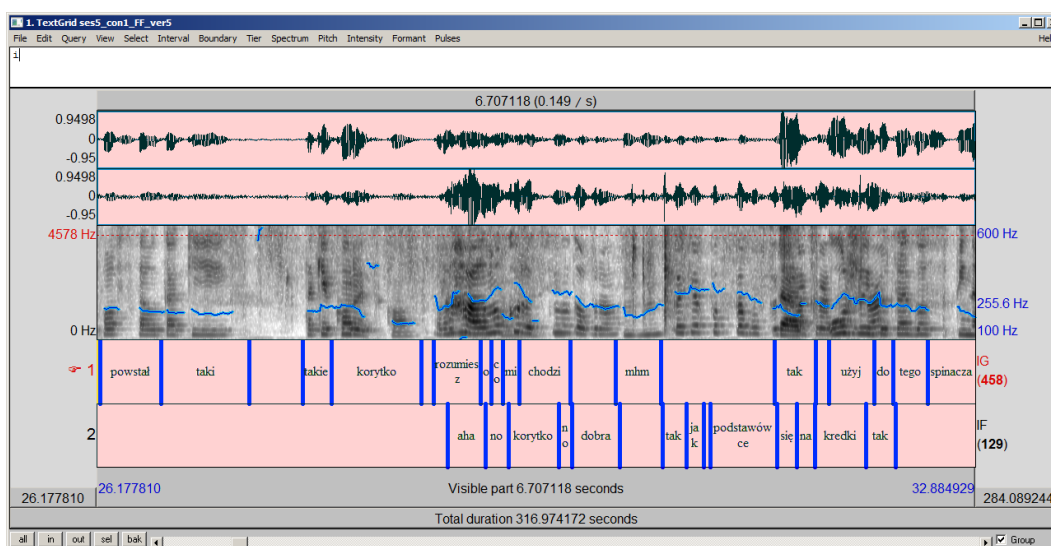
1 Wprowadzenie

Przedmiotem projektu DiaGest2 były badania nad komunikacją multimodalną w dialogach zadaniowych. Analizą objęto cały szereg składników wypowiedzi ustnych (leksykę, wykorzystywane schematy składniowe, prozodię) oraz zachowań niewerbalnych (gestykulację, kierunek spojrzenia, ruchy głowy i tułowia). Dane te posłużyły sformułowaniu modeli wybranych kategorii multimodalnych aktów dialogowych.

W niniejszym tekście przedstawiono analizy pozyskanego materiału leksykalnego i wysunięto hipotezy dotyczące jego własności oraz związków ze specyfiką kontekstu komunikacyjnego i rolami dialogowymi uczestników nagrań.

2 Materiał badawczy

Analizowany tutaj materiał leksykalny pochodzi z dwudziestu sesji dialogowych z udziałem czterdziestu mówców. Zadanie realizowane przez każdą parę mówców polegało na rekonstrukcji papierowej figury przestrzennej. Było ono wykonywane przez dwie grupy (różnych) badanych w dwóch sytuacjach. W pierwszej (wzajemna widzialność: WW) uczestnicy stali naprzeciw siebie i widzieli się nawzajem. Figurę wzorcową widziała osoba nazywana „instruującym” (PI, „podający instrukcje”). Nie widziała jej natomiast druga osoba, nazywana dalej „instruowanym” (OI, „odbierający instrukcje”). PI widział również powierzchnię stołu i materiały niezbędne do odtworzenia figury, które leżały na stole przed OI. W drugim przypadku (ograniczona widzialność: OW) uczestnicy nie widzieli siebie nawzajem. PI widział figurę wzorcową, lecz nie widział osoby OI oraz powierzchni stołu OI i leżących na nim części rekonstruowanej figury. Sesje WW były rejestrowane czterema, OW zaś dwoma kamerami. Fonię zapisywano dodatkowo rejestratorem cyfrowym połączonym z dwoma mikrofonami pojemnościowymi. Czas realizacji zadania był ograniczony do pięciu minut. Nagrania liczą zatem około 100 minut na każdym z dwóch kanałów.



Ryc. 1. Okno programu Praat z warstwami transkrypcji ortograficznej.

Nagrania fonii zostały znormalizowane amplitudowo i przetranskrybowane ortograficznie z wykorzystaniem programu Praat (Boersma, Wenink 2010). Synchroniczna transkrypcja tekstu opierała się na opracowanym przez zespół we wcześniejszych projektach systemie (Karpiński et al. 2008). Na ryc. 1. przedstawiono okno programu Praat z fragmentem transkrypcji ortograficznej. Obejmowała ona nie tylko słowoformy, lecz również inne dźwięki wydawane przez uczestników (w tym jednostki quasi-leksykalne, takie jak „aha” lub „mhm”, pauzy wypełnione, śmiech, westchnienia, okrzyki). Dlatego też w niektórych fragmentach poniższego tekstu mówi się o „jednostkach” lub „tokenach”, nie zaś jedynie o „słowoformach”.

Transkrypty ortograficzne, wykonane w dwóch warstwach (odrębne warstwy dla PI oraz OI), zapisywano w plikach TextGrid, które następnie

importowano do programu ELAN. Z niego eksportowano je w postaci plików tekstowych UTF8, odrębnych dla każdego IG oraz IF. Uzyskano w ten sposób 40 plików tekstowych (dwadzieścia sesji, z każdej dwa pliki).

3 Analizy ilościowe korpusu „origami”

3.1 Ogólna charakterystyka materiału

Materiał w postaci plików tekstowych został poddany analizom ilościowym za pomocą programu AntConc. Zbadano strukturę ilościową materiału i stworzono listy frekwencyjne dla całości oraz poszczególnych części (odrębnie IG i IF zarówno w warunkach wzajemnej widzialności, jak i jej braku). Wygenerowano również listy bigramów.

Zgromadzony materiał stanowi niewielki korpus językowy. Obejmuje łącznie 15.453 tokenów, występujących w 1.719 różnych formach. Część z tych tokenów to symbole specjalne, m.in. tagi zachowań wymienionych w Tabelicy 1. Stosowano je przede wszystkim w przypadku zachowań, które mogły mieć, w ocenie transkrybujących, znaczenie komunikacyjne. Należy odnotować, że tagowanie niektórych z nich było stosunkowo subiektywne (np. „słyszalny oddech”), dlatego też należy ich częstości traktować jako szacunkowe.

Tablica 1. Jednostki nieleksykalne pojawiające się w materiale badawczym
(łącznie liczby wystąpień)

	Jednostka	Objaśnienie	Liczba wystąpień
1	[oddech]	słyszalny oddech	750
2	[pw]	pauza wypełniona (“wypełniacz”, “jęk namysłu”)	553
3	[śmiech]	różne rodzaje śmiechu	159
4	[jęk]	głośne westchnienie, jęknięcie (nie będące jednak “jękiem namysłu”)	69
5	[niezrozumiałe]	niezrozumiały fragment wypowiedzi	60

W materiale pojawiła się też pewna liczba jednostek o stosunkowo niejasnym statusie i znaczeniu często wysoce zależnym od realizacji intonacyjnej i kontekstu. Można do nich zaliczyć m.in. „no” (pełniące w badanym materiale rolę odpowiedzi twierdzącej, wyrazu akceptacji, potwierdzenia, ponaglenia, sygnału gotowości komunikacyjnej, wyrazu zniecierpliwienia; 476 wystąpień), „mhm” (najczęściej pełniące funkcje odpowiedzi twierdzącej, akceptacji, potwierdzenia lub sygnału gotowości komunikacyjnej; 161 wystąpień), „aha” (funkcjonujące jako wyraz akceptacji, zrozumienia, albo odpowiedź twierdząca; 79 wystąpień). Stwierdzono, że jednostka „okey” występowała z reguły w kontekstach, w których mogłaby być substytuowana przez odpowiednio zrealizowane intonacyjnie „tak” (wyrażające zgodę, przyzwolenie, odpowiedź twierdzącą, potwierdzenie, akceptację stanu rzeczy, ale służące również jako marker dyskursowy).

Po odjęciu łącznej liczby wystąpień wymienionych jednostek oraz tagów oznaczających jednostki niezrozumiałe (60 sytuacji, w których przetranskrybowanie wypowiedzi na podstawie odsłuchu okazało się niemożliwe ze względu na zakłócenia akustyczne lub niestaranną artykulację mówcy), w zbiorze pozostało 13.067 realizacji słowoform. W Tablicy 2. wymieniono pierwszych pięćdziesiąt spośród tych jednostek, wraz z ich względnymi częstościami (wobec liczby wszystkich słowoform). Rozkład wydaje się niezwykle charakterystyczny dla języka mówionego (por. Kurcz et al. 1990), a szczególnie dla wypowiedzi dialogowych. Wysokie pozycje zajmują jednostki pełniące kluczowe funkcje interakcyjne (np. „tak” i „nie”, mogące stanowić odpowiedzi na pytania ogólne, jak i reakcje na polecenia, ale również „no” i „czekaj”, mogące pełnić funkcje związane z kierowaniem przebiegiem dialogu lub z zabieraniem głosu), spójniki i inne „słowa funkcyjne”, wykrzyknienia, także kilka form czasowników „mieć” i „być”. Stosunkowo wysokie rangi posiadają niektóre jednostki odnoszące się do artefaktów używanych w realizacji zadania („zapalki”, „spinacze”, „strony”). W Tablicy 2. podano również, w charakterze pewnej płaszczyzny odniesienia, względne częstości jednostek w Narodowym Korpusie Języka Polskiego. Różnice w profilach częstości względnych są wyraźne (zob. Ryc. 2 – ujęto tam pierwszych dwadzieścia pięć haseł z korpusu „origami”). W zestawieniu z listami frekwencyjnymi tworzonymi na podstawie tekstów pisanych, wyraźne różnice w postaci obecności w tych ostatnich jednostek służących budowie zdań złożonych (zaimków względnych lub spójników) są dostrzegalne jednak dopiero w drugiej pięćdziesiątce najczęstszych słowoform (porównanie z korpusem Gazety Wyborczej; Klessa et al. 2010). W dalszych rozważaniach odnoszono się do zbioru jednostek pomniejszonego o jednostki nieleksykalne, wymienione w Tablicy 1. (chyba, że zaznacza się inaczej).

Materiał podzielono na wypowiedzi instruujących i instruowanych, odrębnie dla par widzących i niewidzących się nawzajem. Ilościową strukturę materiału przedstawiono w Tablicy 3. Łatwo zauważyć, że istnieje silna, choć oczekiwana, dysproporcja w liczbie jednostek zrealizowanych w wypowiedziach instruujących i instruowanych. Wynika ona z asymetrycznej natury zadania dialogowego. Odnotujmy, że zadanie – szczególnie w przypadku par widzących się – teoretycznie mogłoby zostać wykonane bez werbalnego udziału instruowanego.

Dostrzegalne są różnice między łącznymi liczbami jednostek w wypowiedziach zarówno PI i OI w obu warunkach (WW i OW). PI w warunkach OW zrealizowali o około 8% więcej jednostek niż w warunkach WW, zaś OI zrealizowali w warunkach OW o około 21% więcej jednostek niż w warunkach WW. Różnice w liczbach bezwzględnych były jednak zbliżone (zob. Tablica 4.). Większą różnicę w liczbach użytych jednostek przez OI można wyjaśnić tym, że (a) w warunkach OW nie mieli oni możliwości obserwowania instruujących i odbierania wizualnych sygnałów sprzężenia zwrotnego, co musieli kompensować werbalnie; (b) nie mogący ich obserwować, a więc pozbawieni możliwości czerpania dodatkowej informacji z gestykulacji instruowani, oczekiwali i wymagali precyzyjniejszego formułowania kierowanych do nich poleceń.

Tablica 2. Pierwszych pięćdziesiąt słowoform z listy frekwencyjnej dla badanego korpusu (podano odsetek wystąpień względem całkowitej liczby słowoform, tj. ok. 13 tys. jednostek; liczba w nawiasie oznacza częstość względną w NKJP policzoną wobec liczby 1,224,800,401 słów)

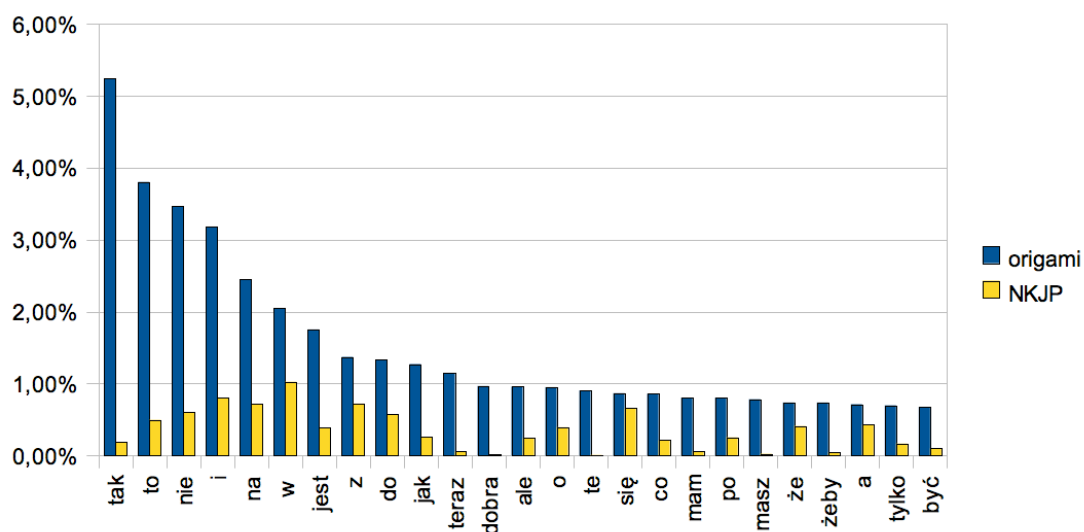
Lp.	Względna częstość w materiale – procenty (nawias: w NKJP)	Jednostka	Lp.	Względna częstość w materiale – procenty (nawias: w NKJP)	Jednostka
1	5,24 (0,18)	tak	26	0,67 (0,10)	ten
2	3,80 (0,49)	to	27	0,65 (0,12)	bo
3	3,47 (0,60)	nie	28	0,65 (0,02)	jakby
4	3,18 (0,80)	i	29	0,64 (0,16)	czy
5	2,45 (0,71)	na	30	0,60 (0,00)	musisz
6	2,05 (1,02)	w	31	0,57 (0,01)	wiesz
7	1,75 (0,39)	jest	32	0,56 (0,15)	są
8	1,37 (0,72)	z	33	0,55 (0,09)	ja
9	1,33 (0,57)	do	34	0,55 (0,04)	wiem
10	1,27 (0,25)	jak	35	0,49 (0,15)	ma
11	1,15 (0,05)	teraz	36	0,49 (0,15)	tego
12	0,96 (0,01)	dobra	37	0,49 (0,00)	zapalki
13	0,96 (0,24)	ale	38	0,46 (0,00)	czekaj
14	0,95 (0,39)	o	39	0,43 (0,05)	takie
15	0,90 (0,01)	te	40	0,42 (0,11)	jeszcze
16	0,86 (0,66)	się	41	0,41 (0,08)	mi
17	0,86 (0,21)	co	42	0,40 (0,04)	czyli
18	0,80 (0,06)	mam	43	0,40 (0,11)	spinacze
19	0,80 (0,25)	po	44	0,39 (0,05)	coś

Lp.	Względna częstość w materiale – procenty (nawias: w NKJP)	Jednostka	Lp.	Względna częstość w materiale – procenty (nawias: w NKJP)	Jednostka
20	0,78 (0,02)	masz	45	0,39 (0,03)	strony
21	0,73 (0,41)	że	46	0,39 (0,03)	taki
22	0,73 (0,04)	żeby	47	0,38 (0,00)	dwie
23	0,71 (0,43)	a	48	0,38 (0,05)	właśnie
24	0,70 (0,16)	tylko	49	0,37 (0,23)	od
25	0,67 (0,10)	być	50	0,37 (0,10)	tej

Warto zwrócić uwagę, że mimo różnic w liczbach zrealizowanych jednostek, leksykony instruujących i instruowanych (rozumiane jako zbiory użytych leksemów) w obu sytuacjach eksperymentalnych były niemal równe co do liczebności. Być może wyjaśnienia należy doszukiwać się w fakcie, iż uczestnicy byli wyraźnie skupieni na realizacji zadania i tylko bardzo nieliczne wypowiedzi wychodziły poza ten zakres tematyczny. Rozmiar leksykonu wydaje się zatem niezwykle silnie zdeterminowany specyfiką zadania. Pewne wsparcie tej tezy w obrębie analizowanego materiału można osiągnąć poprzez porównanie parami list wyrazowych PI_WW - PI_OW oraz OI_WW - OI_BW w celu określenia, czy podobieństwu rozmiaru towarzyszy również podobieństwo profili.

W Tabelicy 4. zawarto listy frekwencyjne pięćdziesięciu najczęstszych jednostek (bez jednostek quasi-leksykalnych, które omówiono odrębnie poniżej). Łatwo dostrzec pewne podobieństwa między zbiorami pięćdziesięciu najczęściej pojawiających się słowoform dla instruujących w warunkach WW i OW. To samo dotyczy instruowanych w obu warunkach realizacji zadania. Wysoka pozycja słowoformy „teraz” w listach frekwencyjnych dla PI wynika ze stosowania jej w instrukcjach jako okolicznika czasu (funkcjonalnie można jej w większości kontekstów uznać za marker dyskursowy, porządkujący tok podawania instrukcji). Oczywiście jest wysoka pozycja słowoformy „musisz” w listach frekwencyjnych dla PI. Bardzo wysokie pozycje w listach dla OI zajmują natomiast leksemy „czekaj” i „poczekaj”, stosowane w wypowiedziach realizujących akty dialogowe kategorii „kierowanie przebiegiem rozmowy”. Jednostka „o” pojawia się częściej u mówców w warunkach WW niż OW. Znowu, hipotetycznie można przyjąć, iż jest to rezultatem częstszego stosowania wyrażen realizujących werbalnie sprzężenie zwrotne, a będących reakcjami na konkretne działania partnera („o właśnie”, „o tak”, „”, i inne).

Michał Szczyszek, Maciej Karpiński: Jakościowa i ilościowa analiza słownictwa w dialogach zadaniowych



Ryc. 2. Częstości względne dla pierwszych dwudziestu pięciu słowoform z korpusu origami oraz ich odpowiedników w NKJP

Tablica 3. Struktura ilościowa wypowiedzi instruujących i instruowanych w warunkach wzajemnej widzialności i jej braku (wartości w nawiasach: bez udziału jednostek nieleksykalnych)

	Widzący się		Niewidzący się	
	PI	OI	PI	OI
Łączna liczba jednostek (nawias: bez jednostek nieleksykalnych)	5013 (4544)	2166 (2044)	5575 (4913)	2699 (2476)
Liczba <u>różnych</u> jednostek (nawias: bez odrzuconych kategorii nieleksykalnych)	902 (897)	518 (513)	894 (889)	530 (535)

Tablica 4. Listy pięćdziesięciu najczęstszych słowoform u PI i OI w warunkach WW i OW (lw – liczba wystąpień)

Lp.	Instruujący (PI)				Instruowani (OI)			
	WW		OW		WW		OW	
	lw	jednost.	lw	jednost.	lw	jednost.	lw	jednost.
1	244	tak	237	tak	126	tak	93	to
2	158	nie	182	i	98	to	78	tak
3	150	i	161	to	74	nie	74	mam
4	145	to	148	nie	44	i	73	nie
5	104	w	137	na	44	jest	64	dobra

Investigationes Linguisticae, vol. XXII

Lp.	Instruujący (PI)				Instruowani (OI)			
	WW		OW		WW		OW	
	lw	jednost.	lw	jednost.	lw	jednost.	lw	jednost.
6	95	na	96	w	43	na	45	na
7	84	jest	83	masz	34	jak	41	w
8	68	do	78	jest	31	ale	40	i
9	66	o	69	z	29	czy	37	jak
10	63	teraz	63	jak	27	o	36	z
11	61	z	58	do	27	w	32	ale
12	49	żeby	57	po	25	teraz	32	co
13	48	te	56	jakby	23	tu	30	czy
14	46	ten	52	teraz	21	być	30	do
15	41	musisz	49	wiesz	20	się	29	ja
16	38	tylko	48	te	19	czekaj	25	bo
17	34	ale	45	się	27	okej	23	czekaj
18	32	jak	44	co	18	do	23	jest
19	31	być	41	żeby	17	dobra	23	okej
20	31	właśnie	36	musisz	17	ma	22	a
21	30	się	35	że	17	mam	21	wiem
22	30	że	34	są	17	wiem	21	że
23	28	jeszcze	33	tylko	16	mi	19	mi
24	27	po	30	a	15	a	18	się
25	26	a	30	ci	15	bo	16	dwie
26	25	dobrze	28	ale	15	ja	14	te
27	25	są	28	spinacze	14	co	13	być
28	23	zapalki	26	bo	14	ten	13	cztery
29	23	znaczy	26	tego	13	tylko	13	po
30	22	co	25	dobra	13	z	13	tego
31	22	musi	25	ma	12	tutaj	13	środką
32	22	ta	24	stronie	11	czyli	12	tam
33	22	tego	24	takie	10	coś	11	czyli
34	22	tej	23	być	10	już	11	jeszcze
35	21	ją	23	kartkę	10	może	11	poczekaj
36	20	dobra	23	prostu	10	poczekaj	11	pudełko
37	20	jakby	23	wiem	10	że	11	zapalki
38	20	taki	22	dwa	9	będzie	10	pionowo

Lp.	Instruujący (PI)				Instruowani (OI)			
	WW		OW		WW		OW	
	lw	jednost.	lw	jednost.	lw	jednost.	lw	jednost.
39	19	bo	22	lewej	9	zapalki	10	rogi
40	19	strony	22	o	8	jeszcze	10	spinacze
41	19	tutaj	22	ten	8	po	10	teraz
42	18	góry	22	tą	8	pół	10	zrobić
43	18	kartkę	21	prawej	8	takiego	9	coś
44	18	takie	21	strony	8	te	9	ma
45	18	zewnątrz	21	zapalki	7	dwie	9	o
46	17	okej	21	zrobić	7	są	9	od
47	17	weź	20	od	7	tam	9	taki
48	16	spinacz	20	okej	7	zewnątrz	9	takie
49	16	środku	20	pudełko	6	chodzi	9	tym
50	15	boki	20	taki	6	musi	9	tą

Dystrybucja jednostek nieleksykalnych i quasi-leksykalnych, których nie ujęto w powyższej tabeli, wykazuje kilka interesujących własności. Dla instruujących, liczba realizacji pauz wypełnionych jest niemal równa w przypadku wzajemnej widzialności i jej braku (odpowiednio, 242 i 234). Z kolei jednostki takie, jak „mhm” i „aha” u instruujących w warunkach WW występowały znacznie częściej niż w warunkach OW. Można to przypisać faktowi, że były one używane jako środek realizacji sprzężenia zwrotnego w reakcji na działania instruowanego, które w warunkach WW instruujący mógł obserwować. Kolejną wartą odnotowania obserwacją jest znacznie częstsze pojawianie się u instruujących jednostki „pw” (etykieta „jęku namysłu”). Ich liczba u instruujących była około pięciokrotnie wyższa niż u instruowanych, co daje istotną różnicę również po uwzględnieniu, że liczba wszystkich zrealizowanych przez instruowanych jednostek była o około połowę mniejsza niż w przypadku instruujących. Wydaje się, że wynikało to ze znacznie większego obciążenia poznawczego, które pojawiało się u instruujących, starających się formułować instrukcje. Czynnikiem szczególnie obciążającym mogła być konieczność zbudowania na podstawie obserwacji gotowego obiektu zbioru procedur zmierzających do jego wykonania, oraz ich precyzyjnej werbalizacji, które pozwoliłyby je realizować instruowanemu bez możliwości oglądania figury.

4 Analiza struktury leksykalnej

4.1 Frekwencja poszczególnych części mowy

Wyrazy tekstowe, będąc realizacjami jednostek leksykalnych, przekazują informację o konkretnej sytuacji, w której zostały użyte bądź do której się odnoszą. W naszym materiale wyrazy tekstowe posłużyły uczestnikom nagrań do opisu sytuacji zadaniowej, która na potrzeby eksperymentu została

zorganizowana. Na podstawie list wyrazów tekstowych można jednakże zrekonstruować strukturę leksykalną analizowanych sytuacji komunikacyjnych, tj. wyabstrahować leksemy, należące do różnych klas części mowy i obliczyć obciążenie realizacyjne poszczególnych klas części mowy w analizowanych sytuacjach komunikacyjnych. Takie ujęcie zagadnienia pozwoli pokazać sposób (mentalny, poznawczy, kognitywny), w jaki zostały ujęte owe sytuacje komunikacyjne przez ich uczestników.

Tablica 5. przedstawia obciążenie realizacyjne w całym analizowanym materiale – wyekscerpowanym z wypowiedzi PI – dla poszczególnych klas części mowy. Są to dane relatywne – odniesione jedynie do całego analizowanego tu korpusu, a więc uwzględniające jedynie charakter konsytuacji – dialogu instruktazowo-zadaniowego zorganizowanego wokół opisanego wyżej w artykule zadania; zatem charakter danych leksykalnych i wniosków na nich opartych jest względny – nie generalny – i można je odnieść tylko do tego typu sytuacji komunikacyjnych i tej konkretnej sytuacji komunikacyjnej, zorganizowanej na potrzeby badań w ramach projektu DiaGest2.

Dominująca pozycja czasownika (V) – patrz Tablica 5. – może świadczyć o tym, że instruktorzy ujmowali sytuację komunikacyjną i zadanie, jakie mieli wykonać, w sposób procesualny, czynnościowy (o semantyce czasownika jako części mowy – i semantyce innych części mowy – zob.: Bańko 2007, Grzegorzczkova 2001, Grzegorzczkova i in. 1986). Wysoka pozycja rzeczownika (S) może sygnalizować, że nadawcy starali się też precyzyjnie kategoryzować opisywaną rzeczywistość, wyodrębniając z tła konkretne przedmioty (obiekty), i – za pomocą czasowników – określać relacje, w jakich się te obiekty względem siebie znalazły. O tym, że bardzo ważne w ujmowaniu i opisywaniu rzeczywistości było określenie relacji zachodzących między obiektami, może świadczyć również wysoka pozycja przysłówków (Adv), przyimków (P) oraz spójników (Conj); w wypadku tych ostatnich istotne jest to, że mają one zdolność łączenia nie tylko wydarzeń (relacji, w których występują obiekty), ale i łączenia obiektów i – co bardzo istotne – wypowiedzeń w jedną większą strukturę komunikacyjną, treściową. Wysoka pozycja zaimków (Pr) zapewne świadczy o wyzyskaniu ich możliwości zastępowania nazw obiektów znajdujących się w opisywanych sytuacjach komunikacyjnych. Odległa, niska pozycja przymiotników (Ad) może świadczyć o tym, że dla nadawców mniej istotne były właściwości obiektów niż właściwości (wyrażane przysłówkami) relacji (procesów, stanów), w jakich znajdowały się opisywane obiekty. Niska pozycja modulantów (Mod) i wykrzyknień (Exc) może świadczyć o tym, że – w procesie komunikacji językowej związanej z wykonywaniem zaprojektowanego na potrzeby eksperymentu zadania – leksemy służące do budowania ramy modalnej, a zatem wyrażania intelektualnego czy emocjonalnego stosunku do nadawcy, odbiorcy, rzeczywistości, czy wypowiedzi (por. Wierzbicka 1971, Jędrzejko 2002), nie są najważniejsze (choć licząc łącznie obciążenie Mod i Exc, okazało się, że wynosi ono ponad 10%; zatem rama modalna stanowi niebagatelny składnik wypowiedzi w tych sytuacjach komunikacyjnych; na uwagę zasługuje tu fakt, że dominuje intelektualny, a nie emocjonalny charakter konstruowanych ram modalnych dla wypowiedzi zadaniowych). Okazało się także, że właściwości kwantytyatywne obiektów i relacji (stanów, procesów) są prawdopodobnie najmniej istotnym składnikiem struktury semantycznej tego typu sytuacji

komunikacyjnych (na co wpływ mogła mieć struktura zadania, które uczestnicy nagrań musieli wykonać).

*Tablica 5. Charakterystyka leksykalna słownictwa użytego przez
instruujących (PI) w obu rodzajach warunków (łącznie): WW i OW*

Klasa części mowy	Odsetek w całości materiału
V	18,06%
S	15,24%
Pr	14,06%
Adv	12%
P	9,95%
Conj	9,48%
Ad	8,05%
Mod	6,68%
Exc	3,36%
Num	3,11%

W Tablicy 5. zawarto – powtórzmy to raz jeszcze – dane „jednostkowe”, uwzględniające tylko materiał pochodzący z nagrań przeprowadzonych w ramach projektu DiaGest2. Zatem informacje o poziomie frekwencji należy traktować tylko relatywnie – względem omawianego tutaj korpusu leksykalnego. Ubezwzględnienie danych, a więc stwierdzenie czy frekwencja danej części mowy jest wysoka czy też nie, można by osiągnąć, porównując dane analizowanego tu korpusu z innymi korpusami leksykalnymi. Taka procedura badawcza pokazałyby bowiem czy np. pozycja przysłówka w analizowanym tu korpusie jest wyjątkowo wysoka, wyjątkowo niska, czy standardowa – w porównaniu z innymi typami sytuacji komunikacyjnych. Językoznawstwo polonistyczne dysponuje kilkoma takimi korpusami ogólnodostępnymi (Kurcz i in. 1990, Korpus Języka Polskiego Wydawnictwa Naukowego PWN, Korpus IPI PAN, PELCRA - korpus referencyjny języka polskiego, Narodowy Korpus Języka Polskiego - Pęzik 2010) i niewiadomą liczbą korpusów nieudostępniowanych publicznie, tworzonych przez poszczególnych badaczy czy zespoły badawcze na własne potrzeby naukowe. Wspomniane tu korpusy ogólnodostępne mają tę właściwość, że notują słownictwo głównie wariantu pisanego polszczyzny, a tylko w nieznacznym stopniu i nie wszystkie – wariantu mówionego. Natomiast analizowany tu korpus gromadzi wyrazy użyte w sytuacjach komunikacyjnych polszczyzny mówionej. Zatem bezpośrednie porównanie korpusu tu analizowanego z korpusami polszczyzny pisanej nie jest możliwe ze względu na zasadniczą

różnicę w charakterze, właściwościach i metodologii obu korpusów, zasadnicza jest także bariera mówioności/pisaności podstawy materiałowej korpusów. Niemniej stworzony w wyniku badań w ramach projektu DiaGest2 korpus jest jednym z pierwszych w językoznawstwie polonistycznym i dla języka polskiego zbiorem leksemów, wyrazów tekstowych użytych w wariacie mówionym polszczyzny (oczywiście są wcześniejsze próby gromadzenia i leksykograficznego opracowywania polszczyzny mówionej, potocznej: *Słownik polszczyzny potocznej* (Anusiewicz, Skawiński 1996) czy powstający od kilku lat krakowski *Słownik polskich leksemów potocznych* (Lubaś 2001-2006) doprowadzony do tomu IV, litery L – lecz albo materiał leksykalny nie był gromadzony metodycznie (Anusiewicz, Skawiński 1996), albo opracowanie leksykograficzne powstaje na podstawie materiału ekscerpowanego z tekstów pisanych, artystycznych, polszczyzny (Lubaś 2001-2006)). Dlatego też, aby odrelatywnić wyniki badań prowadzonych w oparciu o korpus zgromadzony w projekcie DiaGest2, należałoby kontynuować prace nad korpusem lub korpusami wariantu mówionego polszczyzny, w których znalazłyby się leksemy czy wyrazy tekstowe wyekscerpowane z różnego typu sytuacji komunikacyjnych, a nie tylko z dialogu zadaniowego, na podstawie którego powstał analizowany tu korpus.

W Tabelicy 6. przedstawiono różnice w postrzeganiu rzeczywistości, jakie zachodzą w procesie komunikacji międzyludzkiej po wprowadzeniu istotnej zmiennej sytuacyjnej – wzajemnego widzenia się (WW) bądź niewidzenia się (OW) uczestników dialogu. Widoczna jest w takim układzie wyższa dla OW względem Tabelicy 5. (gromadzącej dane o charakterze kontrolnym i porównawczym) frekwencja rzeczownika i niższa frekwencja zaimka. Wynika to prawdopodobnie z faktu, że niewidzący się uczestnicy dialogu muszą precyzyjniej formułować wypowiedź, a zatem muszą precyzyjnie wskazywać na obiekty rzeczywistości, które mają być poddawane zabiegom poznawczym, rekonstrukcyjnym. Możliwe, że skutkiem tego jest także nieco niższa frekwencja (obciążenie realizacyjne) czasowników – nadawcy prawdopodobnie muszą bowiem informować częściej o obiektach, zatem informacje o relacjach stanach, czy procesach między obiektami (w których znajdują się, czy którym podlegają owe obiekty) jest uboższa. Widać to także, gdy się porówna dane uzyskane w warunkach WW z danymi uzyskanymi w OW. W WW wyższą frekwencją niż w OW charakteryzują się: czasownik (tu czasownik ma też wyższą frekwencję niż w Tabelicy 6.), zaimki, przysłówki, wykrzyknienia. Natomiast w warunkach WW niższe wartości niż w OW mają: rzeczowniki, przyimki, spójniki, przymiotniki (pozostałe części mowy mają porównywalne wartości). Może to świadczyć o tym, że w WW możliwe i istotniejsze jest przekazanie informacji o relacji, stanie, procesie obiektu niż o samym obiekcie, który można wskazać za pomocą zaimka (bądź gestu). Natomiast w OW wymagalne jest dokładne określenie obiektu i jego właściwości. Informacja o relacji procesie czy stanie wyrażana czasownikowo jest uboższa, ale uczestnicy nagrań w warunkach OW dokonali swobodnego przeniesienia informowania o relacjach, w jakich znajdują się opisywane przez nich precyzyjnie obiekty, z poziomu leksykalnego (wyrażania czasownikiem) na poziom składniowy, o czym świadczy (tu – na etapie analizy leksykalnej) wyższe niż w WW (i nieco wyższe niż w Tabelicy 6.) obciążenie realizacyjne przyimków i spójników (wniosek ten wymaga oczywiście potwierdzenia w analizach składniowych).

Tablica 6. Charakterystyka leksykalna słownictwa użytego przez
instruujących (PI) w warunkach WW i OW

Widzący się (WW)		Niewidzący się (OW)	
Klasa części mowy	odsetek w materiale	Klasa części mowy	odsetek w materiale
V	18,21%	V	17,93%
Pr	15,25%	S	16,25%
S	14,13%	Pr	12,99%
Adv	12,7%	Adv	11,37%
P	9,23%	P	10,59%
Conj	8,97%	Conj	9,94%
Ad	7,36%	Ad	8,68%
Mod.	6,94%	Mod.	6,45%
Exc	4,24%	Num	3,24%
Num	2,98%	Exc	2,56%

Jeszcze bardziej szczegółowe dane odnośnie opisywanych tu zjawisk leksykalnych przedstawiają Tablice 7. i 8. Zgromadzono w nich dane uzyskane na podstawie analiz wypowiedzi PI w obu warunkach (WW i OW), ale z uwzględnieniem zmiennej płciowej.

Tablica 7. Charakterystyka leksykalna użytego słownictwa przez
instruktorów (PI) w warunkach WW, z podziałem na kobiety i mężczyzn

kobiety		mężczyźni	
V	18,01%	V	18,4%
S	15,26%	Pr	16,09%
Pr	14,42%	S	12,99%
Adv	13,63%	Adv	11,76%
P	10,08%	Conj	8,52%
Conj	9,43%	Mod	8,42%
Ad	6,86%	P	8,38%
Mod	5,46%	Ad	7,86%
Exc	3,59%	Exc	4,89%
Num	3,27%	Num	2,68%

Tablica 8. Charakterystyka leksykalna słownictwa użytego przez
instruujących (PI) w warunkach OW, z podziałem na kobiety i mężczyzn

kobiety		mężczyźni	
V	17,06%	V	18,85%
S	16,82%	S	15,64%
P	11,81%	Pr	14,34%
Pr	11,73%	Adv	11,6%

Adv	11,16%	Conj	9,95%
Ad	9,93%	P	9,3%
Conj	9,93%	Ad	7,34%
Mod	5,66%	Mod	7,3%
Num	3,22%	Num	3,26%
Exc	2,69%	Exc	2,43%

4. 2. Analiza bigramów – frazemów

W Tablicach 9. i 10. przedstawiono bigramy użyte przez PI. Są to struktury dwuwyrazowe, powtarzalne (powtarzające się) w analizowanych sytuacjach komunikacyjnych. Arbitralnie – na potrzeby artykułu – przyjęto konieczny próg powtarzalności się tych struktur na powyżej 6 użyć w całym materiale (czyli 7 i więcej powtórzeń). Te dwuwyrazowe struktury można uznać za przykłady frazemów, a więc konstrukcji językowych, których wyróżnikiem jest powtarzalność w tych samych lub podobnych sytuacjach komunikacyjnych, powtarzalność w tych samych lub podobnych funkcjach (o frazematyce – patrz: Chlebda 2001). Zgromadzone w poniższych tablicach bigramy – frazemy – mogą stanowić przykłady wytwarzania się konstrukcji językowych, mechanizmów komunikacyjnych, stanowiących wyróżnik dla tego typu sytuacji komunikacyjnych, co wygenerowana i analizowana w ramach projektu DiaGest2.

Uwagę zwraca różnica ilościowa między WW a OW: blisko aż o 25% mniej frazemów skonstruowano w WW niż w OW. Istotne jest także i to, że zarówno w WW, jak i w OW owe frazemy budowane są głównie ze spójników, przyimków, modulantów, wykrzykników, zaimków i nieco rzadziej - przysłówków. Niewiele jest tu czasowników, przymiotników i rzeczowników. Można więc założyć, że te frazemy dialogu zadaniowego związanego z zainscenizowaną sytuacją komunikacyjną nie służą raczej przekazywaniu informacji o obiektach i ich właściwościach, o relacjach i ich właściwościach, ale służą raczej podtrzymywaniu kontaktów między uczestnikami dialogu (można przyjąć, że tej konkretnej sytuacji komunikacyjnej ujawnia się czy realizuje w postaci tych frazemów Jakobsonowska funkcja fatyczna), budowania ramy modalnej wypowiedzi (intelektualnej, rzadziej emocjonalnej), a także prób kierowania (zarządzania) komunikacją interpersonalną. Z tych to powodów więcej jest tych frazemów w OW niż w WW.

Tablica 9. Najczęstsze bigramy (występujące przynajmniej siedem razy) w wypowiedziach instruujących (PI) w warunkach wzajemnej widzialności (WW) i ograniczonej widzialności (OW)

Lp.	WW		OW	
	lw	bigram	lw	bigram
1	33	nie nie	31	i teraz
2	30	i teraz	24	wiesz co
3	16	w ten	22	po prostu
4	15	tak tak	21	no to
5	15	to jest	21	tak tak
6	14	do góry	20	nie wiem
7	14	do środka	17	tak jakby
8	14	musi być	17	tak żeby

9	13	o właśnie	14	to jest
10	12	na pół	13	masz to
11	12	no i	12	do środka
12	12	ten sposób	12	lewej stronie
13	11	o tak	12	po lewej
14	9	na zewnątrz	11	i to
15	9	nie tak	11	no i
16	9	nie wiem	11	teraz tak
17	9	tak i	10	ma być
18	9	tak żeby	9	mniej więcej
19	9	w środku	9	nie nie
20	8	i tak	9	po prawej
21	8	i to	8	dwa spinacze
22	8	po prostu	8	i pół
23	8	tylko że	8	no tak
24	7	do tego	8	przed sobą
25	7	i w	7	dwa i
26	7	jeszcze raz	7	i na
27	7	no tak	7	jak to
28	7	o o	7	jest w
29	7	tak samo	7	masz tak
30	7	teraz tak	7	musisz zrobić
31	7	to co	7	na dwa
32	7	to tak	7	na pół
33			7	nie na
34			7	prawej stronie
35			7	tak i
36			7	tak jak
37			7	tak jakbyś
38			7	te rogi
39			7	więc tak
40			7	z tego

5 Wnioski i podsumowanie

Ograniczeniem utrudniającym śmielsze ekstrapolowanie i wnioskowanie, jest niewielki rozmiar badanego korpusu. Z drugiej strony, jest to korpus oparty na rejestracji bardzo szczególnej sytuacji użycia języka, w dużej mierze zdeterminowany specyfiką zadania realizowanego przez mówców. Dlatego też wyciąganie na jego podstawie pewnych wniosków lub choćby stawianie hipotez dotyczących struktury leksyki w podobnych zbiorach wypowiedzi wydaje się uprawnione.

W zgromadzonym materiale zaobserwowano szereg wyraźnych tendencji, wynikających najprawdopodobniej z charakteru realizowanego zadania, lecz również ze specyfiki potocznego języka dialogu i języka mówionego w ogólności. Poniżej podsumowano najistotniejsze z nich.

Odnotowano wysokie częstości względne wielu jednostek związanych z interakcją dialogową, a także niektórych charakterystycznych dla dialogu markerów dyskursowych (np. „teraz”). Na stosunkowo wysokich pozycjach znalazły się również leksemy, których obecność wprost wynikała z charakteru

zadania, na przykład: „zapalki” – jeden z elementów budowanej figury; „musisz” jako składnik wielu poleceń; „masz” jako składnik zdań opisujących sytuację, np. stan figury własnej lub (domniemany) stan figury partnera. „Obok oczywistej różnicy liczby jednostek wypowiedzianych przez instruujących i instruowanych w trakcie trwania zadania, dostrzeżono (a) podobieństwo rozmiarów i profili leksykonów PI_WW i PI_OW, (b) podobieństwo rozmiarów i profili leksykonów OI_WW i OI_OW. Różnice między leksykonami mówców w warunkach WW i OW dotyczyły najczęściej słownictwa związanego ze sferą interaktywności, w tym realizacji sprzężenia zwrotnego. W przypadku warunków OW, to ostatnie musiało opierać się wyłącznie na środkach werbalnych. Być może wpływ na większą liczbę wypowiedzi PI w warunkach OW (w zestawieniu z WW) wpływ miał nie tylko fakt konieczności precyzyjniejszego tłumaczenia, lecz również konieczność realizacji całego procesu sprzężenia zwrotnego drogą werbalną.

Na uwagę zasługuje także fakt, że również struktura leksykalna – analizowanych sytuacji komunikacyjnych – rekonstruowana w trakcie analiz jakościowych – jest zależna od zmiennej WW//OW. Uczestnicy sytuacji OW są bardziej precyzyjni, co przejawia się zwiększoną frekwencją rzeczowników i przymiotników, a więc podawaniem informacji o obiektach i ich właściwościach, niż uczestnicy sytuacji WW, którzy natomiast bardziej zwracają uwagę na relacje między obiektami, korzystając chętniej niż w OW z czasowników, przyimków, spójników i przysłówków. Wnioski uzyskane na podstawie analizy zgromadzonego materiału – korpusu leksykalnego – wymagają potwierdzenia, a zwłaszcza odrelatywnienia, które można by osiągnąć, analizując obszerniejszy korpus, gromadzący dane leksykalne z różnorodnych typów sytuacji komunikacyjnych, zarówno w wariancie mówionym, jak i pisanym języka polskiego.

Wpływ na różnice w wyzyskaniu poszczególnych klas części mowy ma także płeć uczestników analizowanych sytuacji komunikacyjnych. Co prawda kobiety i mężczyźni – niezależnie od sytuacji WW czy OW – równie często używali czasowników, ale kobiety są bardziej precyzyjne niż mężczyźni, częściej używając rzeczowników; natomiast mężczyźni częściej tworzą ramę modalną w swoich wypowiedziach, wyzyskując modulanty.

Bezsprzecznie wpływ na sposoby podtrzymywania kontaktu i na zarządzanie komunikacją miała wprowadzona w eksperymencie zmienna WW//OW, co pokazała analiza bigramów – frazemów. Uczestnicy sytuacji OW, będąc zdani w większej mierze niż uczestnicy WW na kod językowy, wytworzyli znacznie więcej frazemów o funkcji fatycznej, a zatem też pośrednio służących do zarządzania komunikacją językową.

Odnosząc się do całości zgromadzonego materiału (patrząc globalnie na cały korpus – bez podziału na WW i OW), przedstawionego w Tabelicy 6. należy hipotetycznie stwierdzić, że wszyscy PI przyjęli strategię komunikacyjną polegającą na wyróżnieniu w konsytuacji dwóch jej podstawowych składników: a) obiektów i ich cech oraz b) relacji zachodzących między obiektami i właściwościami tych relacji. Tak hipotetycznie rekonstruowana globalna strategia komunikacyjna była – i musiała być – modyfikowana w zależności od udziału w WW czy OW oraz od płci – o czym napisaliśmy powyżej. Jednakże obserwując zgromadzony korpus globalnie, nie można nie zauważyć, że użytkownicy języka polskiego, którzy brali udział w naszych badaniach, podzielili obserwowalną sobie rzeczywistość eksperymentu i dokonali jej reprezentacji intelektualnej właśnie na owe dwa wskazane wyżej składniki. Te

składniki reprezentacji umysłowej komunikowali następnie w poszczególnych wypowiedziach, aktach dialogowych. Warto w tym miejscu odwołać się do koncepcji R. Nisbetta, która zakłada daleko idące różnice w postrzeganiu rzeczywistości zachodzące między przedstawicielami kultury Zachodu i kultur wschodnich (Nisbett analizował głównie kultury Dalekiego Wschodu), a polegające na tym, że ludzie Zachodu są nastawieni na obiekt, szczególnie, wyodrębnianie z tła, natomiast ludzie Wschodu – łatwiej opisują relacje, tło, kontekst (Nisbett 2009). Dane zgromadzone w analizowanym korpusie są oczywiście daleko niewystarczające, aby móc dokonać generalnego porównania sposobów poznawania i kategoryzowania rzeczywistości dokonywanych przez użytkowników polszczyzny, a więc mieszkańców Europy Środkowej, pominiętej w badaniach Nisbetta, z danymi, które uzyskał autor *Geografii myślenia*. Dane z analizowanego korpusu powstałego w projekcie DiaGest2 mogą stanowić jedynie przyczynek do dalszych badań w tym zakresie – badań potwierdzających, modyfikujących czy obalających hipotezę, że użytkownicy języka polskiego w sytuacji zadaniowej, w dialogu zadaniowym na pierwszy plan w kategoryzowaniu poznawanej rzeczywistości zadaniowej i komunikowaniu o tej rzeczywistości wysuwają relację (stan, czynność, proces – komunikowane przez czasownik), którą starają się wyraźnie doprecyzować wyodrębniając precyzyjnie obiekty i ich właściwości. Oczywiście ta hipoteza wymaga dalszych badań przeprowadzanych na materiale zarówno zdaniowych sytuacji komunikacyjnych, jak i – zwłaszcza – innych typów sytuacji komunikacyjnych.

Projekt, w ramach którego prowadzono opisane badania, skupia się na stworzeniu modeli wybranych multimodalnych modeli aktów dialogowych. Mają one obejmować również wskazówki, co do zawartości werbalnej wypowiedzi. Przeprowadzone tutaj analizy zostaną poszerzone o badania leksykalnej zawartości realizacji wybranych kategorii aktów dialogowych.

Wydaje się, iż jednym z kolejnych kroków ku precyzyjnierzemu opisowi i interpretacji struktury słownictwa w badanym tutaj materiale powinno być również skonfrontowanie przedstawionych profili częstościowych słownictwa z analogicznymi profilami uzyskanymi dla podobnych korpusów. Pożądane byłoby wykorzystanie korpusów powstałych na kanwie zarejestrowanych w analogicznych warunkach zadań dialogowych innych typów. Zgodnie z wiedzą autorów, takie korpusy dla polszczyzny jeszcze nie powstały (choć liczbę korpusów dialogowych rośnie; wśród nich warto odnotować korpus LUNA, gdyż jego dialogi można w dużej mierze uznać za zadaniowe (Mykowiecka et al. 2010)). Niezbędnym w najbliższej przyszłości krokiem wydaje się również bardziej szczegółowe otagowanie całości analizowanego materiału, z uwzględnieniem znaczeń poszczególnych jednostek. Pozwoliłoby to na wykonanie precyzyjniejszych list frekwencyjnych oraz konfrontację ich z podobnie opisanymi listami z innych korpusów.

Podziękowanie

*Niniejsze badania zostały zrealizowane w ramach projektu
N N104 010337, finansowanego przez MNiSW.*

Bibliografia

- Anusiewicz, J., Skawiński, J. 1996. *Słownik polszczyzny potocznej*, Warszawa -Wrocław.
- Bańko, M. 2007. *Wykłady z polskiej fleksji*. Warszawa.
- Boersma, P., Wenink, 2010. *Praat 3.9: Doing Phonetics by Computer*. (computer software)
- Chlebda, W. 2001. Frazematyka. w: J. Bartmiński (red.) *Współczesny język polski*. Lublin: s. 335-342
- Grzegorzczkova, R. 2001. *Wprowadzenie do semantyki językoznawczej*. Warszawa.
- Grzegorzczkova, R., Laskowski, R., Wróbel, H. (red.). 1986. *Gramatyka współczesnego języka polskiego*, t. II: *Morfologia*. Warszawa: Wydawnictwo PWN.
- Jędrzejko, E. 2002. Modalność - kategoria opisowa stylistyki i tekstologii lingwistycznej. w: Z. Krążyńska i Z. Zagórski (red.) *Poznańskie Spotkania Językoznawcze*, t. IX, Poznań: Wydawnictwo PTPN, s. 57-67.
- Karpiński, M., Jarmolowicz-Nowikow, E., Malisz, Z., Szczyszek, M., E., Juszczak, K. 2008. Rejestracja, transkrypcja i tagowanie mowy oraz gestów w narracji dzieci i dorosłych. (Recording, transcription and tagging of speech and gestures in child and adult narration). Poznań, *Investigationes Linguisticae*, Volume XVI: 83-98.
- Klessa, K., Karpiński, M., Bałdys, O., Demenko, G. 2009. Speechlabs ASR. Polish Lexical Database for Speech Technology: Design and Architecture. *Speech and Language Technology*, vol 11.
- Kurcz, I., Lewicki, A., Sambor, J., Szafran, K., Woronczak, J. 1990. *Słownik frekwencyjny współczesnej polszczyzny*. t. I–II. Kraków
- Lubaś, W. (red.). 2001-2006. *Słownik polskich leksemów potocznych*. t. I-IV (A-L), Kraków: Wydawnictwo Naukowe DWN.
- Mykowiecka A., K. Głowińska and J. Rabeiga-Wiśniewska 2010. Domain-related Annotation of Polish Spoken Dialogue Corpus LUNA.PL, *Proceedings of LREC 2010 Conference*, Malta.
- Nisbett, R.E. 2009. *Geografia myślenia*. Przekład E. Wojtych. Sopot: Wydawnictwo Smak Słowa.
- Pęzik, P. 2010. Providing corpus feedback for translators with the PELCRA Search Engine for NKJP. In: *Practical Applications in Language and Computers 2009*.
- Wierzbicka, A. 1971. Kategoria semantyczna „ramy” modalnej. w: *Sesja Naukowa Międzynarodowej Komisji Budowy Gramatycznej Języków Słowiańskich*. Wrocław: s. 179-183.

Łącza internetowe

[Korpus IPI PAN](#)

[Korpus Języka Polskiego Wydawnictwa Naukowego PWN](#)

[Narodowy Korpus Języka Polskiego](#)

[PELCRA - korpus referencyjny języka polskiego](#)

[Program ELAN](#)

[Program PRAAT](#)

[Projekt LUNA](#)