

Anna Seretny  
Uniwersytet Jagielloński, Kraków  
aseretny@poczta.onet.pl

## ZASÓB LEKSYKALNY OBCOJĘZYCZNYCH UŻYTKOWNIKÓW POLSZCZYZNY W ŚWIETLE BADAŃ IŁOŚCIOWYCH<sup>1</sup>

**Słowa kluczowe:** kompetencja leksykalna, wskaźniki bogactwa leksykalnego, profil leksykalny wypowiedzi, uczący się języka polskiego

**Keywords:** lexical competence, indices of lexical richness, lexical profile, Polish language learners

### 1. Wprowadzenie

*Kompetencja leksykalna* definiowana jest w glottodydaktyce jako znajomość słów i umiejętność posługiwania się nimi w aktach komunikacji. Definicja ta, aczkolwiek jasna i zrozumiała, nie jest precyzyjna. Nie mogłaby więc w tym kształcie służyć jako podstawa badań (ilościowych i/lub jakościowych), nie wiadomo bowiem, jak należy rozumieć „słowo” i co jest przejawem jego „znajomości”. Nie wynika z niej, jaki charakter ma „umiejętność posługiwania się słowami”, nic nie mówi również o rodzaju i stopniu złożoności aktów komunikacji, w których użytkownik ma uczestniczyć. Konieczne było zatem dookreślenie wymienionych pojęć i wskazanie zachodzących między nimi relacji oraz ich operacjonalizacja (przełożenie na mierzalne wskaźniki).

---

<sup>1</sup> Wybór tematu artykułu ma podwójny, aczkolwiek oczywisty związek z osobą Jubilata, który zajmował się polszczyzną osób dwujęzycznych, a słownictwu i jego badaniu poświęcił wiele prac, podkreślając w nich konieczność prowadzenia nie tylko analiz jakościowych, lecz także ilościowych. Zainteresowania te stały się dla mnie źródłem inspiracji, której owocem jest niniejsza praca.

Dzięki takim zabiegom możliwe stało się wskazanie różnych aspektów kompetencji leksykalnej użytkownika, którymi są:

- wielkość zasobu słownikowego (*breadth of vocabulary knowledge* – znajomość słownictwa „wszerz”), określana liczbą znanych mu jednostek, którymi są zarówno pojedyncze wyrazy, jak i wieloelementowe ciągi o różnym stopniu zespolenia (zob. ESOKJ: 102);
- stopień znajomości leksyki tworzącej zasób (*depth of vocabulary knowledge* – znajomość słownictwa „w głąb”), który można wyrazić liczbą znanych użytkownikowi komponentów znajomości poszczególnych jednostek<sup>2</sup> (zob. Richards 1976; Nation 2001; Seretny 2011);
- poziom znajomości jednostek zasobu – receptywny, uzewnętrzniający się jedynie możliwością rozpoznawania ich znaczeń i funkcji w tekstach mówionych i pisanych, lub produktywny, przejawiający się również umiejętnością posługiwania się nimi (*receptive vs productive vocabulary knowledge*);
- dostępność jednostek słownika (*accessibility of lexical items*), czyli tempo, w jakim wydobywane są z pamięci w aktach komunikacji.

Można było podjąć także poszukiwania bardziej trafnych i wiarygodnych sposobów ich opisywania i badania.

## 2. Kompetencja leksykalna a różni użytkownicy polszczyzny

Wielkość zasobu słownikowego, stopień znajomości tworzących go jednostek oraz poziom ich dostępności są czynnikami silnie warunkującymi kształt i przebieg aktu komunikacji, a często wręcz decydującymi o jego powodzeniu, dlatego też rozwój kompetencji leksykalnej jest niezwykle ważny zarówno w procesie przyswajania języka pierwszego, jak i uczenia się następnego. Opanowywanie słownictwa w obu wypadkach ma charakter równie złożony, a zanim zapis nowo poznawanych jednostek w słowniku mentalnym zostanie utrwalony i pozwoli użytkownikom na sprawne posługiwanie się nimi, przechodzi przez wiele stadiów: począwszy od pierwszego śladu umożliwiającego im jedynie rozpoznawanie wyrazów w określonych kontekstach, po pełny rejestr cech, otwierający przed nimi szerokie spektrum działań produktywnych. Etapy pośrednie bywają liczne, a ich sekwencja, zwłaszcza w wypadku uczących się, może mieć charakter addytywny lub progresywno-regresywny (zob. Seretny 2011).

Zasób leksykalny rodzimych użytkowników języka (UR<sup>3</sup>), znajomość słownictwa „wszerz”, bywa bardzo zróżnicowany, przy czym głównymi determinantami

2 To jest znajomości znaczenia jednostek, ich form morfologicznych, łączliwości składniowo-semantycznej, konotacji, odcieni znaczeniowych. Użycie musi być bowiem zarówno poprawne, jak i fortunate komunikacyjnie.

3 W niniejszym tekście skróty UR, UC i UPP oznaczają, odpowiednio, użytkowników rodzimych, cudzoziemskich i polskiego pochodzenia.

jego obszerności są poziom wykształcenia oraz stopień czytania. Wielkość słownika poszczególnych osób nie jest też stała, lecz zmienia się w czasie, przede wszystkim ze względu na otwarty charakter systemu leksykalnego, a także intensywność kontaktu z jego poszczególnymi obszarami<sup>4</sup>. Fluktuacje dotyczą jednak przede wszystkim peryferii, a nie centrum zasobu (*core vocabulary*). Znakomitą większość zasobu tworzą wyrazy biernie znane użytkownikom, stosunkowo niewielka zaś jego część jest w stałym użyciu lub może być bez trudności aktywowana. Czynne posługiwanie się jednostkami leksykalnymi wymaga ich wysokiej dostępności, tj. szybkiego przetwarzania, wiąże się też z koniecznością ich dobrej znajomości „w głąb”; bierne natomiast polega głównie na łączeniu formy jednostek (fonicznej i graficznej) ze znaczeniem, które powinno przebiegać na tyle sprawnie, by nie utrudniało ani nie hamowało procesu rozumienia.

Rodzimi użytkownicy poszerzają zasobność swojego słownika w dwojaki sposób: naturalny, w trakcie interakcji z otoczeniem, oraz kierunkowany – w wyniku nauki szkolnej (zob. tabela 1). Szacuje się, że drogą pierwszą w dzieciństwie poznają około 4000–6000 wyrazów, tak bowiem ocenia się zasób leksykalny sześciu-, siedmiolatków (Kielar-Turska 1989). Następnie rocznie dodają do swojego „leksykonu” około 1500–2000 słów, nie tyle jednak ucząc się ich, ile za ich pomocą poznając świat w szkole. Liczba wyrazów znana dorosłym jest, jak już wspomniano, bardzo zróżnicowana. Zazwyczaj przyjmuje się, że ich słownik mentalny liczy jednak co najmniej kilkanaście tysięcy jednostek, w niektórych przypadkach sięgając kilkudziesięciu lub nawet kilkuset tysięcy.

Przyrostu warstwy leksykalnej cudzoziemców uczących się języka (UC) nie daje się ująć w ramy czasowo-liczbowe ze względu na wielość czynników warunkujących intensywność procesu uczenia się<sup>5</sup>. Dlatego też częściej mówi się o wielkości zasobu słownikowego oczekiwanej na konkretnym poziomie zaawansowania niż o tym, w jakim czasie można go osiągnąć (Seretny 2011). Dla osób poznających język w warunkach sztucznych (zob. tabela 1) problemem jest nie tylko liczba jednostek, których znaczenie muszą przyswoić, by móc swobodnie porozumiewać się w języku docelowym (znajomość „wszerz”), lecz przede wszystkim opanowanie wielu aspektów ich znajomości (znajomość „w głąb”). Rzadko więc osiągają poziom kompetencji leksykalnej porównywalny z rodzimymi wykształconymi użytkownikami języka.

Między wspomnianymi wyżej grupami, traktowanymi często jako przeciwstawne, lokuje się zbiorowość trzecia, dla której dany kod, w naszym przypadku polszczyzna, jedynie po części jest językiem ojczystym, po części zaś pozostaje obcym. Należą do niej emigranci i ich potomkowie (UPP) zanurzeni na co dzień w mowie

4 Część wyrazów poznawanych w toku nauki szkolnej staje się po czasie „pustymi etykietami”. W wielu wypadkach użytkownik potrafi je co najwyżej przyporządkować do określonej dziedziny.

5 Są to, m.in., wiek uczących się, poziom ich motywacji, intensywność nauki (liczba godzin w tygodniu), odległość typologiczna i genetyczna dzieląca ich J1 od poznawanego.

i kulturze kraju osiedlenia, a po polsku porozumiewający się jedynie w kręgu rodzinnym, sąsiedzkim lub co najwyżej w lokalnym środowisku etnicznym (zob. tabela 1). Ich swój/nieswój polski nazywa się językiem odziedziczonym (JOD)<sup>6</sup>. Stopień opanowania JOD przez użytkowników może być różny, od doskonałego po szczątkowy, przy czym za główne czynniki utrudniające i/lub hamujące jego rozwój, a niekiedy przyczyniające się do częściowego zaniku, uważa się konieczność opanowania języka kraju osiedlenia (Polinsky, Kagan 2007; Montrul 2008) oraz ograniczony w JOD, niekiedy do zera, wymiar edukacji szkolnej (Lipińska, Seretny 2012a). Przełączenie się na kod docelowy przy braku nauki obniża bowiem poziom aktywności w JOD, skutkując niedorozwojem kompetencji dyskursywnej i tekstowej oraz bardzo podstawowymi umiejętnościami w zakresie czytania i pisania, zwłaszcza jeśli porównać je z posiadanymi przez rodzimych użytkowników w podobnym wieku. Ograniczona znajomość sprawności manualno-wzrokowych, określana przez Tse (2002) mianem piśmienności utajonej (*latent literacy*), i tak sytuuje ich jednak wyżej od rozpoczynających naukę cudzoziemców.

Tabela 1. Źródła danych leksykalnych dla trzech grup użytkowników polszczyzny<sup>7</sup>.

	UR	UPP	UC
RODZINA	+	+	-
OTOCZENIE SPOŁECZNE – SKALA MAKRO	+	-	-
OTOCZENIE SPOŁECZNE – SKALA MIKRO (LOKALNA SPOŁECZNOŚĆ)	+	?	-
SYSTEM EDUKACJI	+	-	-
LEKCJE JĘZYKA	-	?	+

Osoby polskiego pochodzenia posługują się językiem dość sprawnie (idiomatycznie w rozumieniu Sinclair'a 1991), zwłaszcza gdy przedmiotem ich wypowiedzi są kwestie związane z codziennością, należące do podstawowych kręgów tematycznych<sup>8</sup>. Język cudzoziemców, którzy są na podobnym poziomie biegłości (zazwyczaj jest to co najmniej B1/B2), wydaje się natomiast leksykalnie bardziej nieporadny, „chropawy”, brzmi zdecydowanie mniej naturalnie, choć z niezweryfikowanych danych wynika, że cechuje go nieco większe bogactwo środków wyrazu. Wolniejszy w przypadku UC jest też dostęp do zasobów słownika mentalnego.

6 Zagadnienia związane z językiem odziedziczonym (*heritage language*) zyskały na znaczeniu od czasu pojawienia się publikacji Valdés (1997, 2000) i Fishmana (2001).

7 Wszystkie tabele w tekście są autorskie.

8 Ekspresja osób dorosłych w języku odziedziczonym pozostaje jednak często na poziomie nastolatka. Ich wypowiedzi są proste, powierzchowne, tematycznie ograniczone do doświadczeń codziennych. W języku kraju osiedlenia natomiast te same jednostki potrafią wyrażać się precyzyjnie i elokwentnie (Miodunka 2003).

### 3. Zasób leksykalny cudzoziemców (UC) i użytkowników języka odziedziczonych (UPP) w świetle badań ilościowych

#### 3.1. Cel badań

Niniejszy artykuł przedstawia wyniki badań pilotażowych, które miały na celu ustalenie, czy sposób poznawania języka – *naturalny* lub *mieszany* (*naturalny* i *szkolny*) w przypadku UPP vs *szkolny* w przypadku UC – wpływa na bogactwo leksykalne ich wypowiedzi pisemnych, innymi słowy, czy w zasobach leksykalnych UPP i UC widoczne są różnice, a jeśli tak, to jakiego rodzaju.

W analizach wykorzystane zostały wskaźniki ilościowe, gdyż po pierwsze, uzyskane na ich podstawie wyniki umożliwiają wnioskowanie o natężeniu zjawiska w całej populacji; po drugie zaś, ujawnienie u badanych potencjalnych różnic co do znajomości „wszerz” daje prawo przypuszczać, iż odmiennie ukształtowana jest u nich także sieć powiązań między tworzącymi zasób jednostkami (znajomość „w głąb”), w obszarze kompetencji leksykalnej różnice jakościowe są bowiem zawsze pochodną ilościowych, a nie odwrotnie (zob. Vermeer 2004).

#### 3.2. Materiał badawczy

Były nim wypowiedzi pisemne przedstawicieli obu grup będących na poziomie średnim ogólnym B2. Pozyskano go z archiwów Państwowej Komisji Poświadczania Znajomości Języka Polskiego jako Obcego. Było to 40 prac (podzielonych na dwie równoliczne grupy) z sesji egzaminacyjnych odbywających się w roku 2013. Ich dobór miał charakter losowy, a czynnikiem decydującym o włączeniu do korpusu badawczego był wybór przez zdających konkretnego zestawu egzaminacyjnego (nr 2). Dzięki temu wszystkie badane wypowiedzi:

- poruszały ten sam temat („Reklama telewizyjna w naszym życiu”),
- były ujęte w ramy jednakowej formy (wybór tematu narzucał konieczność napisania rozprawki),
- miały podobną długość (około 260 wyrazów, +/- 10%),
- powstały w identycznym przedziale czasu (zdający mieli łącznie 90 minut na napisanie 300 wyrazów<sup>9</sup>).

Założono, że jednostki zawarte w wypowiedziach stanowią aktywny słownik użytkowników, do którego mają szybki i elastyczny dostęp.

Autorami prac byli cudzoziemcy z 9 krajów świata oraz 20 użytkowników języka odziedziczonych z USA (zob. tabela 2) w wieku od 17 do 22 lat. W obu grupach nieznacznie przeważała płeć żeńska.

9 W ramach zestawu egzaminacyjnego zdający mieli jeszcze napisać ogłoszenie liczące 40 wyrazów (+/- 10%).

Tabela 2. Narodowość autorów prac (z uwzględnieniem sposobu poznawania języka).

Narodowość badanych					
Cudzoziemcy (UC)			Uczący się polskiego pochodzenia (UPP)		
Kraj	Liczba	%	Kraj	Liczba	%
Ukraina	6	30	USA	20	100
Niemcy	4	20			
Słowenia	3	15			
Białoruś	2	10			
Sri Lanka	2	10			
Węgry	1	5			
Francja	1	5			
Libia	1	5			

### 3.3. Metodologia badań

Do ilościowej charakterystyki tekstów napisanych przez obie grupy wykorzystane zostały następujące, omówione poniżej, wskaźniki: zróżnicowania leksykalnego (WZL), gęstości (WGL) oraz poprawności leksykalnej (WPL), a także wskaźnik nominalności tekstu (WNT). Zbadano też leksykalny profil ich wypowiedzi (LPW), który jest zespołem wskaźników ilościowo-jakościowych (Laufer, Nation 1995).

Określenie WZL polega na wskazaniu zależności zachodzącej między wyrazami stanowiącymi *słownik* tekstu a wszystkimi w nim użytymi (*tekstem*). Im wskaźnik ten jest wyższy, tym wypowiedź jest bogatsza pod względem leksykalnym<sup>10</sup>. WGL z kolei pokazuje, w jakim stopniu tekst został nasycony wyrazami autosemantycznymi. Większa jego wartość uznawana jest za przejaw rozleglejszego słownika. WGL wypowiedzi pisemnych musi przekraczać 50% (zob. O'Loughlin 1997), w przeciwnym razie uznaje się je za wtórnie pisane (pierwotnie mówione), co dodatkowo potwierdzić można, obliczając WNT, czyli stosunek rzeczowników do czasowników<sup>11</sup>. Za pomocą WPL stwierdza się, ile wyrazów tekstu zostało prawidłowo użytych, obrazując w ten sposób stopień kontroli autora nad leksykalną poprawnością przekazu. Dzięki leksykalnemu profilowaniu wypowiedzi (LPW) dowiadujemy się zaś, do ja-

10 WZL nie pozwala jednak stwierdzić, jakiego rodzaju wyrazy zostały użyte, czy trudniejsze, rzadziej używane, czy też bardziej podstawowe – częściej spotykane. Jest także bardzo czuły na długość tekstu (im tekst krótszy, tym zazwyczaj wskaźnik wyższy).

11 W tekstach pisanych kategorie nominalne powinny zdecydowanie dominować nad werbalnymi (Kamińska-Szmaj 1990: 134). Pisarek (1972: 74–75) zauważa, że polszczyzna pisana jest stylem typu *wi* w odróżnieniu od polszczyzny mówionej, w której spójnik *i* ma rangę wyższą niż przyimek *w* (zob. Zgólkowa 1983: 167).

kiej kategorii należą jednostki użyte w tekście<sup>12</sup>: czy są to słowa o bardzo wysokiej, średniej czy też niskiej frekwencji (obecność tych ostatnich traktuje się jako wskaźnik większej biegłości językowej).

Wykorzystanie narzędzi statystycznej analizy leksykalnej zawartości tekstów miało posłużyć weryfikacji następujących hipotez badawczych:

- A. Wskaźniki zróżnicowania leksykalnego i gęstości leksykalnej prac osób polskiego pochodzenia oraz prac obcokrajowców będą różne.
- B. Prace osób polskiego pochodzenia cechować będzie niższy wskaźnik nominalności, charakterystyczny dla języka mówionego (zakłada się, że język pisany ich tekstów może być zapisem mówionego).
- C. Prace cudzoziemców będą miały niższy wskaźnik poprawności leksykalnej.
- D. Profil leksykalny wypowiedzi osób polskiego pochodzenia będzie różny od profilu prac uczących się obcokrajowców w zakresie wykorzystania słownictwa najczęściej i rzadko używanego.

### 3.4. Opracowanie materiału badawczego

Zastosowanie narzędzi statystycznych wymagało uprzedniego opracowania materiału. Wyrazy tekstowe każdej pracy sprowadzono więc do form podstawowych, usuwając jednocześnie z korpusu badawczego wszystkie nazwy własne (wskazywałyby one na użycie słownictwa bardzo rzadkiego) oraz liczebniki (jako nieobecne na listach frekwencyjnych). Powstały zasób dzielono na dwie grupy: do pierwszej włączano jednostki autosemantyczne, tj. rzeczowniki, czasowniki, przymiotniki oraz przysłowki, do drugiej – synsemantyczne. Za błędne uznawano (zob. m.in. Dąbrowska, Pasięka 2006):

- niepoprawnie utworzone derywaty słowotwórcze (wybór niewłaściwej podstawy, niewłaściwego formantu);
- wyrazy nieistniejące w języku polskim;
- wyrazy użyte w niewłaściwym znaczeniu;
- wyrazy konotujące treści inne niż zamierzano;
- niepoprawnie utworzone kolokacje (będące rezultatem naruszenia łączliwości semantycznej);
- wyrazy z innego piętra taksonomicznego (zbyt ogólne lub zbyt szczegółowe) lub z niewłaściwego rejestru.

Jako poprawne z leksykalnego punktu widzenia traktowano natomiast jednostki leksykalne, w obrębie których pojawiały się niewielkie odstępstwa od normy gramatycznej i/lub ortograficznej.

---

12 Profil ten jest niezależny od tematu wypowiedzi, jej długości, języka pierwszego autorów, a jego wyniki zmieniają się wraz z rosnącym poziomem ich biegłości językowej (zob. Meara 2005; Laufer 2005; Edwards, Collins 2011).

#### 4. Analiza wyników badań

Podjęte badania miały służyć ilościowemu określeniu stopnia podobieństwa lub rozbieżności między próbkami wypowiedzi obu grup użytkowników polszczyzny. Wszystkie wskaźniki zostały więc wykorzystane nie jako mierniki wartości estetycznych tekstów, lecz jako narzędzia umożliwiające opis struktury statystycznej badanego materiału.

Analizie poddano prace o łącznej długości 10 807 słowoform. Teksty UPP liczyły ich 5376, UC – 5431 (zob. tabela 3). Średnia długość tekstu to – dla grupy pierwszej 269, dla drugiej – 271 słowoform. Mimo ścisłych wymogów egzaminacyjnych poszczególne wypowiedzi liczyły od 184 aż do 316 wyrazów tekstowych, przy czym zdecydowanie większa jednorodność cechowała prace cudzoziemców (mniejsza wartość odchylenia standardowego). Wielkość słownika (liczba jednostek) obcokrajowców była wyraźnie większa niż u osób polskiego pochodzenia, tekst natomiast (liczba słowoform) – był w zasadzie porównywalny (zob. tabela 4).

Tabela 3. Wielkość korpusu badawczego (w słowoformach).

	Liczba wypowiedzi	Całkowita długość wypowiedzi
UPP	20	5376
UC	20	5431
RAZEM	40	10807

Tabela 4. Średnia długość wypowiedzi w słowoformach, średnia wielkość słownika (*type*) i tekstu (*token*).

	Średnia długość wypowiedzi		Średnia wielkość słownika		Średnia długość tekstu	
	MEAN	SD	MEAN	SD	MEAN	SD
UPP	268,80	40,40	127,55	16,89	246,05	38,58
UC	271,50	31,30	137,20	15,21	249,80	28,50

Unikanie powtórzeń w wypowiedzi (wskaźnik WZL) wynikające z zastosowania szerszego zakresu słownictwa nie tylko znacząco wpływa na jej walory stylistyczne (przejaw wrażliwości na słowo), świadczy także o lepszej znajomości leksyki „wszerz” przez użytkownika. Brak umiejętności różnicowania może zaś być wynikiem mniejszego zasobu, jaki ma on do dyspozycji, nieumiejętnego wykorzystania dostępnych mu środków (słabe wyczulenie na słowo) lub niskiego stopnia ich automatyzowania<sup>13</sup>. Najbardziej zróżnicowane leksykalnie wypowiedzi (zob. tabela 5)

13 Gdy trzeba szybko reagować, a z taką sytuacją mamy do czynienia na egzaminie, użytkownik sięga zazwyczaj do tych danych, do których ma bezwiedny, bezpośredni dostęp.



zostały napisane przez cudzoziemców (nr 6 – WZL 71,2%; nr 9 – WZL 72,5%), najmniej – przez osobę polskiego pochodzenia (nr 13 – WZL 41,6%). Zadowolająca wartość WZL to przynajmniej 50% (zob. Read 2000). Nie uzyskało jej 12 prac, w tym 5 autorstwa cudzoziemców (nr 5, 10, 11, 19, 20) oraz 7 napisanych przez osoby polskiego pochodzenia (nr 6, 11, 12, 13, 14, 15, 20). Średni WZL dla obu grup różni się nieznacznie na korzyść obcokrajowców. Niewielkie odchylenia standardowe w badanych przypadkach wskazują na stosunkowo dużą jednorodność obu populacji pod względem badanej cechy (zob. tabela 6).

Wskaźnik gęstości leksykalnej (WGL) pokazuje stopień nasycenia wypowiedzi treścią (jej nośnikami są wyrazy autosemantyczne). Dla tekstów pisanych, jak już wspomniano, powinien on przybierać wartość powyżej 50%. Nie uzyskała jej tylko jedna praca użytkownika JOD (nr 11). Średni WGL jest ponownie nieco wyższy dla wypowiedzi obcokrajowców (zob. tabela 6), a odchylenie standardowe przyjmuje podobną, niewysoką wartość, dając świadectwo ich niewielkiego wewnętrznego zróżnicowania.

Teksty pisane w języku polskim charakteryzuje wysoki stopień nominalności, osiągający wartość 2 i większą (zob. Kamińska-Szmaj 1990: 136), mówione natomiast niski, około 0,87 (zob. Zgółkowa 1983: 171). Średnie wskaźniki nominalności (WNT) wypowiedzi obu grup nie różnią się znacząco, choć rozstęp wyników u badanych polskiego pochodzenia jest zdecydowanie mniejszy (od 0,74 do 1,5) od rezultatów osiągniętych przez cudzoziemców (od 0,55 do 2,07). Znamienny wydaje się też fakt, że żadna praca użytkownika JOD nie osiągnęła poziomu 2, a taki odnotowany został w przypadku czterech wypowiedzi cudzoziemców<sup>14</sup>.

Niewielkie różnice widać także w zakresie wskaźnika poprawności leksykalnej (WPL), który przybiera stosunkowo wysokie wartości<sup>15</sup>, co oznacza, że wszyscy badani dość sprawnie używają znanego im słownictwa.

---

14 Wstępna analiza jakościowa, wykraczająca poza ramy niniejszego opracowania, wskazuje na nadużywanie przez osoby polskiego pochodzenia czasowników *być* oraz *mieć*.

15 Dotyczy on wyrazów, których użyto trafnie, posługując się ich znaczeniem, adekwatnie do kontekstu, bez naruszenia norm łączliwości leksykalnej. Mogły natomiast zostać błędnie odmienione lub zastosowane w wadliwej konstrukcji składniowej.

Tabela 5. Profil leksykalny wypowiedzi UPP i UC wraz z ich parametrami leksykalnymi.

Praca	Profil leksykalny wypowiedzi (w %)				WSKAŹNIKI (w %)				
	1-1000	1001-2000	2001-3000	3000-...	WZL	WGL	WPL	WNT	
UC	1	78,8	16,1	3,38	1,69	50,6	57,5	100,0	1,16
	2	72,0	15,6	4,5	7,1	57,6	58,0	86,7	1,09
	3	79,2	12,3	2,3	6,1	54,6	65,5	92,2	1,86
	4	77,1	10,0	6,4	6,4	56,4	63,7	94,2	1,28
	5	79,8	12,9	3,6	3,6	49,2	56,7	95,3	0,86
	6	61,8	24,4	6,9	6,9	71,2	68,4	89,3	2,07
	7	71,2	18,9	3,9	5,9	53,9	62,6	89,8	1,19
	8	74,2	16,1	4,8	4,8	55,6	65,4	95,2	1,36
	9	63,5	18,5	9,3	8,6	72,5	65,3	84,2	1,73
UC	10	78,0	13,4	5,6	2,8	49,4	58,2	93,1	1,26
	11	73,5	13,2	6,6	6,6	47,8	61,6	93,3	1,69
	12	69,7	16,5	7,9	5,7	55,1	52,3	92,6	1,21
	13	76,7	12,4	3,9	7,0	53,0	69,1	91,6	2,03
	14	81,8	8,7	3,5	5,8	57,3	67,1	95,8	1,37
	15	71,1	14,8	7,4	6,6	57,6	65,4	93,2	1,12
	16	73,2	14,5	6,1	6,1	53,0	64,4	93,2	1,00
	17	68,6	16,4	7,4	8,9	60,9	66,8	90,9	2,00
	18	61,7	19,1	6,6	12,6	60,9	66,8	87,2	2,00
	19	78,1	15,6	4,7	1,6	47,7	57,4	89,9	0,55
20	77,7	12,5	2,6	7,1	44,1	57,8	92,7	1,08	
UPP	1	79,8	12,8	5,3	6,3	51,3	63,3	92,4	1,36
	2	80,8	9,6	4	5,6	52,9	60,5	87,4	1,05
	3	72,1	15,5	6,2	6,2	53,7	57,5	96,7	1,30
	4	74,4	11,5	6,2	7,9	51,3	55,4	89,4	1,02
	5	76,3	11,8	4,1	7,6	55,5	62,1	97,3	0,94
	6	66,9	19,0	8,2	5,7	49,2	64,6	87,2	1,01
	7	70,1	13,1	7,3	9,4	56,6	64,0	85,2	1,48
	8	57,6	20,6	10,8	10,8	56,7	66,0	88,0	0,83
	9	74,1	13,9	4,2	7,7	59,0	61,9	89,3	1,51
	10	78,7	14,7	2,2	4,4	55,7	59,4	87,7	1,12
	11	73,6	12,8	9,3	4,2	49,6	41,4	93,3	1,44
	12	71,9	14,9	10,2	2,8	48,6	62,7	89,7	1,12

Praca	Profil leksykalny wypowiedzi (w %)				WSKAŹNIKI (w %)				
	1-1000	1001-2000	2001-3000	3000-...	WZL	WGL	WPL	WNT	
UPP	13	83,2	9,0	4,1	3,4	41,6	56,6	93,4	0,92
	14	80,3	8,2	4,1	7,3	44,3	57,4	97,5	0,74
	15	75,0	17,2	6,0	1,7	44,9	60,8	90,8	1,42
	16	69,8	17,1	4,8	8,2	50,3	58,2	93,2	1,37
	17	73,2	16,9	1,4	8,4	66,0	61,8	94,5	1,14
	18	76,3	10,8	3,4	9,4	54,0	58,0	87,8	1,50
	19	77,7	14,0	2,9	5,2	57,4	57,0	94	1,29
	20	80,5	11,0	4,2	4,2	46,2	58,8	92,05	0,98

Tabela 6. Parametry leksykalne wypowiedzi HLL i FLL.

	Parametry leksykalne							
	WZL		WGL		WPL		WNT	
	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
UC	55,40	6,97	62,50	4,61	92,02	3,51	1,36	0,42
UPP	52,24	5,61	59,31	5,02	91,34	3,54	1,17	0,23

Tabela 7. Profil leksykalny wypowiedzi HLL i FLL.

	Profil leksykalny wypowiedzi							
	1-1000		1001-2000		2001-3000		3000-...	
	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
UC	73,4	5,80	15,10	3,42	5,37	1,87	6,10	2,44
UPP	74,6	5,68	13,72	3,24	5,44	2,52	6,32	2,35

leksykalny profil wypowiedzi (LPW) obu grup użytkowników polszczyzny jest również bardzo zbliżony<sup>16</sup> (zob. tabela 7, tabela 5) i mieści się w przedziale „późnego B2 – wczesnego C1” (Seretny 2011: 285). Słownictwo z pierwszego przedziału częstości w podobnym zakresie pokrywa ich teksty, przy czym wypowiedzi o najniższym (praca nr 8 – 57,6%) i najwyższym (praca nr 13 – 83,2%) stopniu nasycenia wyrazami o wysokiej frekwencji (przedział 1-1000) zostały napisane przez osoby polskiego pochodzenia. W przypadku wyrazów z drugiego i trzeciego przedziału częstości (zob. tabela 7) zarówno średnie obu grup, jak i odchylenie standardowe przybierają analogiczne niemal wartości.

16 W analizach wykorzystano listy frekwencyjne w układzie gronowym (zob. Seretny 2011).

Tabela 8. Parametry leksykalne wypowiedzi HLL i FLL – wyniki testu t-Studenta.

	Parametry leksykalne					
	WZL		WGL		WPL	
	MEAN	SD	MEAN	SD	MEAN	SD
UC	55,40	6,97	62,50	4,61	92,02	3,51
UPP	52,24	5,61	59,31	5,02	91,34	3,54
T – TEST						
	t = 1,52 df 38 ( $\alpha = 0,05$ ) p < 0,01		t = 1,3 df 38 ( $\alpha = 0,05$ ) p < 0,01		t = 0,62 df 38 ( $\alpha = 0,05$ ) p < 0,01	

Tabela 9. Profil leksykalny wypowiedzi HLL i FLL – wyniki testu t-Studenta.

	Profil leksykalny wypowiedzi							
	1-1000		1001-2000		2001-3000		3000-...	
	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
UC	73,4	5,80	15,10	3,42	5,37	1,87	6,10	2,44
UPP	74,6	5,68	13,72	3,24	5,44	2,52	6,32	2,35
T – TEST								
	t = -0,42 df 38 ( $\alpha = 0,05$ ) p < 0,01		t = 1,29 df 38 ( $\alpha = 0,05$ ) p < 0,01		t = -0,22 df 38 ( $\alpha = 0,05$ ) p < 0,01		t = -0,20 df 38 ( $\alpha = 0,05$ ) p < 0,01	

Teksty pisane przez cudzoziemców i osoby polskiego pochodzenia mają zatem podobne parametry leksykalne (WZL, WGL, WNT, WPL), zbliżony jest również leksykalny profil ich wypowiedzi (LWP). Potwierdzają to wartości średnich oraz niewielkie odchylenia standardowe. Celem upewnienia się, że różnice między średnimi wynikami obu grup rzeczywiście nie są istotne, przeprowadzony został test t-Studenta dla grup równolicznych o równych wariancjach (przy stopniu prawdopodobieństwa  $\alpha = 0,05$ , stopniu swobody  $df = 38$  i przedziale ufności  $p < 0,01$ ). Wykazał on jednoznacznie, że w żadnym z badanych przypadków **nie mamy do czynienia** z różnicą statystycznie istotną (zob. tabele 8, 9). Próbkę wypowiedzi obu grup użytkowników polszczyzny, **wbrew przewidywaniom**, nie okazały się zatem wyraźnie inne tak pod względem ilościowym (WZL, WGL), jak i ilościowo-jakościowym (WNT, LWP). Żadna zatem z postawionych hipotez badawczych dotycząca zasobności słownika użytkowników JOD i cudzoziemców nie znalazła potwierdzenia w wynikach analiz próbek ich wypowiedzi pisemnych.

## 5. Podsumowanie

Prowadzone badania miały charakter pilotażowy. Ich celem było ustalenie, czy wypowiedzi pisemne użytkowników języka odziedziczonego oraz uczących się polszczyzny obcokrajowców mają podobne parametry leksykalne. Zakładano bowiem, że inny sposób opanowywania języka (*naturalny* lub *mieszany* vs *szkolny*) oraz inne zdiagnozowane obszary trudności u przedstawicieli obu grup (zob. Lipińska, Seretny 2012b) znajdą swoje odbicie w jakości ich pisemnej produkcji językowej.

Czynniki te nie mają jednak takiego znaczenia, jakie im w założeniach przypisywano. Wskaźniki bogactwa słownikowego oraz profile leksykalne wypowiedzi przedstawicieli obu grup są zbliżone. Jedyną ważną różnicę odnotowano na płaszczyźnie słownika (*types*), który w wypadku cudzoziemców jest obszerniejszy niż u osób polskiego pochodzenia (zob. tabela 5). Potwierdza to obserwację znaną z praktyki dydaktycznej, że ci drudzy dość opornie rozbudowują swój zasób leksykalny, najczęściej nie przykładając wagi do intensywnej pracy nad jego poszerzaniem.

Powodów konieczności odrzucenia większości postawionych hipotez badawczych może być kilka. Po pierwsze, próbki wypowiedzi pochodziły z egzaminów certyfikacyjnych, do których uczący się polskiego pochodzenia musieli się przygotowywać<sup>17</sup>, a zatem w czasie poprzedzającym pomiar w ich biografii językowej szkolny sposób opanowywania umiejętności językowych zdominował naturalny<sup>18</sup>. Po drugie, sposób interpretowania umiejętności leksykalnych jest zawsze determinowany przez kontekst (Read 2000). Temat wypowiedzi, wpisujący się w prywatną sferę życia, pozwolił im na większą swobodę<sup>19</sup>, obcokrajowcom dał zaś możliwość sięgania do dobrze utrwalonych pokładów słownictwa (sfera prywatna jest także tą, do której odwołuje się większość materiałów dydaktycznych). Powodem trzecim mogło być zaś zbyt małe zróżnicowanie etniczno-językowe grupy cudzoziemskiej, wśród której połowę stanowiły osoby pochodzenia słowiańskiego. Słowianie stanowią specyficzną podzbiorowość uczących się (zob. Seretny 2011), sytuującą się ze względu na typologiczną i genetyczną bliskość ich języków i polszczyzny wyraźnie **między** osobami polskiego pochodzenia i „klasycznymi” obcokrajowcami.

Postawione hipotezy mogłyby zostać w pełni zweryfikowane po przeprowadzeniu badań na większej i bardziej zróżnicowanej populacji, w których komponent ilościowy zostałby uzupełniony jakościowym (analiza poszczególnych losowo wybranych

---

17 Zapoznając się ze *Standardami wymagań egzaminacyjnych, Przykładowymi testami* oraz innymi materiałami dydaktycznymi itp.

18 Wszyscy zdający egzamin byli uczniami polskich szkół uzupełniających w Chicago i Nowym Jorku, dzięki czemu ich język domowy, który zazwyczaj cechuje niewielkie bogactwo środków wyrazu, stał się kodem bardziej rozwiniętym.

19 Zagadnienia związane z codziennością, a za takie można uznać rolę reklam telewizyjnych w naszym życiu, są dla osób polskiego pochodzenia bardziej „oswojone”, łatwiej więc podlegają werbalizacji.

próbek wypowiedzi). Uzyskane wyniki niewątpliwie przyczyniłyby się bowiem do pogłębienia naszej wiedzy o kompetencji leksykalnej różnych użytkowników polszczyzny. Ta zaś pozwoliłaby na usprawnianie procesu kształcenia językowego i jego właściwe ukierunkowywanie.

## Literatura

- CUMMINS J., 1982, *Bilingualism and Minority Language Children*, Toronto.
- DĄBROWSKA A., PASIEKA M., 2006, *Błąd językowy – niedostatek kompetencji, luka w sprawności czy niewłaściwa strategia*, [w:] A. Seretny, E. Lipińska (red.), *Sprawności przede wszystkim*, Kraków, s. 15–35.
- EDWARDS R., COLLINS L., 2011, *Lexical frequency profile and Zipf's law*, „Language Learning” nr 61(1), s. 1–30.
- ESOKJ: *Europejski system opisu kształcenia językowego: uczenie się, nauczanie, ocenianie*, Warszawa 2003.
- FISHMAN J., 2001, *300-plus years of heritage language education in the United States*, [w:] J.K. Peyton, D.A. Ranard, S. McGinnis (red.), *Heritage languages in America: Preserving a national resource*, Washington, s. 81–89.
- KAMIŃSKA-SZMAJ I., 1990, *Różnice leksykalne między stylami funkcjonalnymi polszczyzny pisanej. Analiza statystyczna na materiale słownika frekwencyjnego*, Wrocław.
- KIELAR-TURSKA M., 1989, *Mowa dziecka. Słowo i tekst*, Kraków.
- LAUFER B., 2005, *Lexical frequency profiles: from Monte Carlo to the real world: a response to Meara (2005)*, „Applied Linguistics” nr 26(4), s. 582–588.
- LAUFER B., NATION P., 1995, *Vocabulary size and use: lexical richness in L2 written production*, „Applied Linguistics” nr 16(3), s. 307–322.
- LIPIŃSKA E., SERETNY A., 2012a, *Między językiem ojczystym a obcym. Nauczanie i uczenie się języka odziedziczonego w chicagowskiej diasporze polonijnej*, Kraków.
- LIPIŃSKA E., SERETNY A., 2012b, *Uczący się języka polskiego jako obcego/drugiego w środowisku endolingwalnym – ich umiejętności i potrzeby*, „Kwartalnik Polonicum” nr 12(1), s. 2–10.
- MEARA P., 2005, *Lexical frequency profiles: a Monte Carlo analysis*, „Applied Linguistics” nr 26(1), s. 32–47.
- MIODUNKA W., 2003, *Bilingwizm polsko-portugalski w Brazylii*, Kraków.
- MONTRUL S., 2008, *Second language acquisition welcomes the heritage language learner: opportunities of a new field*, „Second Language Research” nr 24(4), s. 487–506.
- NATION I.S.P., 2001, *Learning Vocabulary in Another Language*, Cambridge.
- O'LOUGHLIN K., *Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test*, „Language Testing” nr 12, s. 217–237.
- PISAREK W., 1972, *Frekwencja wyrazów w prasie. Wiadomości – komentarze – reportaże*, Kraków.
- POLINSKY M., KAGAN O., 2007, *Heritage languages: in the “wild” and in the classroom*, „Compass of Language and Linguistics” nr 1(5), s. 368–395.
- READ J., 2000, *Assessing Vocabulary*, Cambridge.
- RICHARDS J., 1976, *The role of vocabulary teaching*, „TESOL Quarterly” nr 1, s. 77–89.

- SERETNY A., 2011, *Kompetencja leksykalna uczących się języka polskiego jako obcego w świetle badań ilościowych*, Kraków.
- SINCLAIRE J., 1991, *Corpus, Concordance, Collocation*, Oxford.
- TSE L., 2002, *Heritage language literacy: A study of U.S. biliterates*, „Language, Culture, and Curriculum” nr 14, s. 256–268.
- VALDÉS G., 1997, *The teaching of Spanish to bilingual Spanish-speaking students: Outstanding issues and unanswered questions*, [w:] M.C. Colombi, F.X. Alarcón (red.), *La Enseñanza del Español a Hispanohablantes*, Boston, s. 93–101.
- VALDÉS G., 2000, *The teaching of heritage languages: An introduction for Slavic-teaching professionals*, [w:] O. Kagan, B. Rifkin (red.), *Teaching and Learning Slavic Languages and Cultures*, Bloomington, s. 375–403.
- VERMEER A., 2004, *The relation between lexical richness and vocabulary size in Dutch L1 and L2 Children*, [w:] P. Bogaards, B. Laufer (red.), *Vocabulary in a Second Language*, Amsterdam–Philadelphia, s. 173–190.
- ZGÓŁKOWA H., 1983, *Ilościowa charakterystyka polszczyzny*, „Studia Polonistyczne” nr 10, s. 163–175.

## **Assessing lexical resources of Polish language learners using lexical frequency profile**

### **Summary**

Polish is learnt by two types of learners; namely, heritage language learners (HLLs) and foreign language learners (FLLs). HLLs are of Polish origin and have gained their partial knowledge of language in a natural environment unlike FLLs who learn Polish in a formal institutional setting and have no Polish roots whatsoever.

Lexical resources of HLLs and FLLs and their access to mental lexicon language data, even when their overall level of language proficiency is similar, seem to differ. HLLs speak more fluently, particularly when talking about everyday matters, and their production sounds more native like, being more formulaic. FLLs on the other hand, when speaking, “look for words” with more difficulty and sentences they build often sound odd.

The aim of the project is to find out whether the lexical resources of HLLs are really, qualitatively and quantitatively, different to that of FLLs and if so, to what extent. In order to estimate the vocabulary size of the two groups samples of HLLs’ and FLLs’ writing at B2 level were analyzed with the use of various indices such as lexical variation, density, sophistication and Lexical Frequency Profile.