

Anna Bączkowska

Uniwersytet Kazimierza Wielkiego, Wydział Humanistyczny

Korpusowa analiza dyskursu związanego z tematyką imigracji w brytyjskiej prasie opiniotwórczej

Corpus-based analysis of immigration discourse in British broadsheet press

Abstract: The aim of the paper is to show the applicability of corpus-driven methodology to media discourse analysis. Theoretically, the study assumes that the semantic meaning of a lexical item is emergent and distributed, i.e. it resides in the word under investigation as well as the collocations with which it co-occurs. Methodologically, the analysis is primarily quantitative and relies on automatic data extraction and statistical calculations. The data used for the study were gleaned from a corpus of the British press, which contains articles coming from three broadsheet newspapers representing both a conservative and a left-wing political stance. The paper demonstrates how a corpus-driven methodology can be successfully employed in media studies, in particular in press discourse analysis.

Słowa kluczowe: dyskurs imigracyjny, prasa opiniotwórcza, analiza korpusowa

Keywords: immigration discourse, broadsheet newspapers, corpus-driven analysis

Wstęp

Niniejszy artykuł przybliży problem imigracji przedstawiony w brytyjskiej prasie opiniotwórczej (ang. *broadsheet press*). Badanie dotyczy artykułów opublikowanych przez kilka gazet opiniotwórczych w latach 1993, 2005 i 2010 reprezentujących różne poglądy polityczne. Opisana poniżej analiza języka mediów wykorzystuje metodologię stosowaną w językoznawstwie (korpusowym). Analiza dyskursu oparta na korpusach językowych jest co prawda znana na gruncie polskim, ale jej zastosowanie w badaniach medioznawczych wydaje się być pewnym *novum*. Niniejszy artykuł pokazuje, w jaki sposób można wykorzystać korpusy gazet w badaniach medioznawczych i po jakie narzędzia dostępne w analizie korpusowej można w tym celu sięgnąć. Celem artykułu jest zatem nie tylko przybliżenie problemu imigracji, podejmowanego w opiniotwórczej prasie brytyjskiej, ale również omówienie zastosowania metodologii opartej na korpusach w badaniach medioznawczych.

Podstawy teoretyczno-metodologiczne

Korpus językowy to znaczący pod względem wielkości zbiór danych określonego języka, które są magazynowane w formie elek-

tronicznej i analizowane za pomocą narzędzi informatycznych, tj. programów komputerowych służących do przeszukiwania i ekstrakcji danych. Automatyczne przeszukiwanie danych możliwe jest po ich wcześniejszej anotacji morfoskładniowej (obejmującej segmentację, lematyzację, parsowanie itp.). Korpus jest nie tylko anotowany, ale też ustrukturyzowany, tzn. zawiera dane wybrane i zapisane według jakiegoś wcześniej określonego kryterium (np. roku wydania gazet czy reprezentowanych przez nie poglądów politycznych, okresu historycznego dzieł literackich itp.). Nie jest to zatem baza danych składająca się z przypadkowego zbioru słów.

Badania językoznawcze na gruncie polskim korzystają z korpusów językowych od około dwóch dekad. Od 2005 r. do dyspozycji językoznawców (jak i zwykłych użytkowników Internetu) jest nieodpłatny dostęp do Narodowego Korpusu Języka Polskiego¹ (NKJP), który powstał z połączenia kilku korpusów tworzonych przez badaczy z ośrodków naukowych w Polsce (z Warszawy, Łodzi i Krakowa). Korpus udostępniony jest online przez PWN z podstawowymi narzędziami ułatwiającymi przeszukiwanie i analizę danych korpusowych. Bardziej zaawansowane statystycznie narzędzia do analizy korpusów oferuje kolokator Hask². Na gruncie badań anglosaskich badania językoznawcze oparte na danych korpusowych prowadzone są już od lat sześćdziesiątych XX wieku. Od początku ich istnienia korpusy wykorzystywane są do analizy leksyki, frazeologii i składni, znacznie rzadziej natomiast do analizy dyskursu. Możliwości, jakie daje korpus w analizie dyskursu zarówno w badaniach językoznawczych, jak i tych na pograniczu badań społecznych, opisał dość szczegółowo w 2006 r. brytyjski językoznawca Paul Baker³. Analiza korpusowa w badaniach społecznych jest nowością w Polsce i, jak już wspomniano, niniejszy artykuł ma na celu przybliżyć tę metodologię badaczom języka mediów.

Teoretycznie rzecz ujmując, badania korpusowe mają często charakter emergentny i dystrybucyjny⁴. Emergentność rozumiana jest jako sposób analizy polegający na wnioskowaniu w oparciu o obserwacje wyłaniające się w trakcie samej analizy, bez założeń wstępnych przyjętych przez badacza przed rozpoczęciem badania.

¹ A. Przepiórkowski, M. Bańko, R. Górski, B. Lewandowska-Tomaszczyk. (red.). *Narodowy Korpus Języka Polskiego*, PWN, Warszawa 2012.

² P. Pezik, „Paradygmat dystrybucyjny w badaniach frazeologicznych. Powtarzalność, reprodukcja, idiomatyzacja”, [w:] *Metodologie językoznawstwa. Ewolucja języka. Ewolucja teorii językoznawczych*, red. naukowa P. Stalmaszczyk, Łódź 2013.

³ P. Baker, *Using Corpora in Discourse Analysis*, Londyn 2006.

⁴ B. Lewandowska-Tomaszczyk, *Podstawy językoznawstwa korpusowego*, Łódź 2005.

Analizując na przykład znaczenie semantyczne jakiegoś słowa w korpusie sprawdza się konkretne konteksty (zwykle są to setki lub tysiące przykładów), w których ono występuje i na tej podstawie dokonuje się abstrahowania części składowych bardziej ogólnej definicji. Procedura ta (od szczegółu do ogółu) zakłada, że znaczenie wyłania się w trakcie analizy, czyli jest emergentne i dynamiczne (w zależności od typu korpusu definicja tego samego słowa może podlegać pewnym modyfikacjom i wykazywać różnice). Takie podejście do znaczenia (znane w literaturze anglosaskiej jako *corpus-driven*, wprowadzone przez Tognini-Bonelli w 2001 roku⁵, lub jako technika probabilistyczna, zob. Bączkowska 2003⁶), zakłada również, że znaczenie leksykalne i propozycjonalne nie tkwi tylko w statycznej definicji pojedynczego słowa, lecz w kontekście, w którym występuje, obejmującym grupy słów. Kontekst bowiem profiluje znaczenie słowa kluczowego, pozwala je doprecyzować i wskazuje na jego preferencje semantyczne. Słowa współwystępujące ze słowem kluczowym (kolokaty) nie muszą bezpośrednio występować w jego sąsiedztwie (tj. bezpośrednio je poprzedzać lub po nim następować). Mogą one pojawiać się w pewnej odległości (zwykle do pięciu słów w lewym lub prawym kontekście), co wskazuje zatem na ich rozproszenie. O ile są one rekurentne (powtarzające się w korpusie), stają się miarodajne i mogą stanowić podstawę dla konceptualizacji i definiowania znaczenia słowa kluczowego. Jeśli znaczenie słowa kluczowego wyłania się z kontekstu, w którym występuje, to znaczenie to jest rozproszone (ang. *distributed*), czyli ma charakter dystrybucyjny. Istotny dla analizy opartej na korpusach jest również fakt, że znaczenie ustala się na podstawie danych autentycznych (tzw. uzus językowy). Teksty zawarte w korpusach pozyskuje się z tekstów opublikowanych, np. literatury czy prasy, lub z transkrypcji teksów mówionych. Pozwala to na analizę języka odzwierciedlającą faktyczny „stan rzeczy”, a nie jedynie na odwoływanie się do statycznego znaczenia leksykograficznego, które jest przewidywalne i uniwersalne.

Alternatywna procedura analizy opartej na korpusie zakłada, że dane korpusowe stanowią materiał pozwalający jedynie zweryfikować wcześniejsze hipotezy czy teorie. W literaturze anglojęzycznej takie podejście znane jest jako *corpus-based*. W niniejszym artykule przyjęto jednak podejście *corpus-driven*, które zakłada, że znaczenie leksykalne i propozycjonalne jest dynamiczne i dystrybucyjne, bowiem znaczenie słowa kluczowego jest zależne od znaczeń tego słowa oraz

⁵ T. Tognini-Bonelli, *Corpus Linguistics at Work*, John Benjamins, Amsterdam 2001.

⁶ A. Bączkowska. *Wybrane techniki wykorzystania korpusów językowych w nauczaniu języka angielskiego*, *Języki Obce w Szkole* 3, 2003, s. 64–76.

wyrazów, z którymi badane słowo współwystępuje. Kryterium dys-trybucyjne jest zatem podstawą określania znaczenia.

Korpus SiBol/Port (SBP)

Korpus SiBol/Port (SBP), na którym oparta jest niniejsza analiza, powstał w wyniku współpracy dwóch uniwersytetów we Włoszech (w Sienie i Bolonii) oraz uniwersytetu w Wielkiej Brytanii (w Portsmouth)⁷. Korpus liczy 327.025.669 segmentów i składa się z 787 artykułów opublikowanych w brytyjskiej prasie jakościowej (*broadsheet newspapers*) w latach 1993, 2005 i 2010. Korpus zawiera artykuły pochodzące z następujących gazet: „The Times” (T), „The Guardian” (G), „The Daily Telegraph” (DT), „The Sunday Times” (ST), „The Sunday Telegraph” (STg). W przeważającej większości zatem materiał korpusowy reprezentuje poglądy prawicowe. Gazety o nachyleniu konserwatywnym występujące w korpusie SBP to „The Times” i jego wariant weekendowy „The Sunday Times” oraz „The Daily Telegraph” i jego wariant weekendowy „The Sunday Telegraph”. Jediną gazetą znaną z popierania poglądów lewicowych i prospołecznych jest „The Guardian”. Korpus jest dość dobrze zbalansowany pod względem struktury przy uwzględnieniu kryterium daty publikacji artykułów 27,2% tekstów pochodzi z 1992 r., 39,5% stanowią teksty z 2005 r., natomiast 33,3% badanych źródeł to teksty opublikowane w 2010 r. Nieco gorzej kształtuje się struktura korpusu ze względu na tytuł gazety. Najwięcej artykułów pochodzi z gazety „The Times” (34,4%), „The Daily Telegraph” stanowi 28% korpusu, „The Guardian” 27%, „The Sunday Times” 7,6%, „The Sunday Telegraph” natomiast 3,2%.

Przedstawiona poniżej analiza przeprowadzona została z wykorzystaniem systemu *The Sketch Engine*⁸ (SK) i kolokatora *Hask*⁹ (dostępne online). Zastosowano oferowane przez nie komputerowe narzędzia statystyczne zaprojektowane dla analizy języka. Ograniczeniem, ale i jednocześnie zaletą przedstawionych w niniejszym artykule badań jest wielkość korpusu. Ograniczenie wynika z tego, że korpus zawiera dane obrazujące jedynie trzy lata publikacji. Z drugiej strony jednak (o ile mi wiadomo) jest to jedyny korpus dostępny komercyjnie, który zawiera dane z więcej niż jednej gazety, co pozwala dokonać pewnych analiz porównawczych. Warto też wspomnieć, że korpus ten, mimo iż zawiera artykuły zebrane tylko

⁷ Autorami tego projektu są Alan Scott Partington (Bologna University), John Morley (Siena University), Anna Marchi (Lancaster University) i Charlotte Taylor (University of Sussex).

⁸ Strona internetowa: the.sketchengine.co.uk

⁹ Strona internetowa: pelcra.pl/hask_pl/home

z trzech lat, jest trzy razy większy od standardowego korpusu tzw. drugiej generacji – British National Corpus (liczącego 98 tysięcy segmentów; BNC). Jako korpus referencyjny BNC jest punktem odniesienia dla innych korpusów języka angielskiego. Badania oparte na tych danych, mimo iż nie są pełne, to jednak mogą stanowić podstawę do dalszych korpusowych badań porównawczych.

Analiza

Jednym z podstawowych narzędzi w metodologii językoznawstwa korpusowego jest lista frekwencyjna (*word list*). Charakterystyka frekwencyjna pozwala zaobserwować, ile jest wystąpień słowa kluczowego w całym badanym tekście (korpusie różnych gazet, korpusie pojedynczej gazety, czy korpusie pojedynczej gazety w konkretnym roku wydania) w porównaniu ze wszystkimi innymi słowami. Słowo występujące często pojawi się na początku listy frekwencyjnej, a występujące rzadziej pojawi się na końcu listy. Drugim istotnym parametrem jest liczba wystąpień słowa kluczowego w korpusie w konkretnych kontekstach (liczba konkordancji) oraz ich wartość standaryzowana, na przykład w obliczeniu na milion słów. Te dwa parametry w odniesieniu do słowa kluczowego *immigrant* ilustruje poniższa tabela 1. Ogółem, w całym korpusie SBP jest 6800 wystąpień słowa *immigrant*, tj. 17,54 słowa (segmentu) na milion, po ustandaryzowaniu danych.

Tabela 1. Liczba wystąpień i frekwencja słowa *immigrant*

<i>Immigrant</i>	1993 liczba wystąpień (na milion)/frekwencja	2005 liczba wystąpień (na milion)/frekwencja	2010 liczba wystąpień (na milion)/frekwencja
The Guardian	823(2,12)/518	1076(2,78)/605	817(2,11)/496
„The Times”	473(1,22)/282	998(2,57)/560	762(1,77)/438
The Daily Telegraph	366(0,94)/194	763(2,00)/463	722(1,86)/461

Jak wskazują powyższe dane z lat 1993 i 2010 częstość występowania słowa *immigrant* w gazecie „The Guardian” jest porównywalna. Natomiast dane z 2005 r. dla tej gazety mają najwyższą wartość w całym korpusie SBT. Kwestia imigracji w Wielkiej Brytanii jest problemem społecznym, a „The Guardian”, jako przykład prasy o poglądach centrolewicowych i kojarzony głównie z brytyjską Partią Pracy, donosi o aktualnych sprawach społecznych, co może wyjaśniać dane statystyczne przedstawione powyżej. Warto wspomnieć przy okazji, że między 1997 a 2010 rokiem rządy w Wielkiej Brytanii sprawowane były przez Partię Pracy, która prowadziła łagodną politykę imigracyjną (m.in. otworzyła granice dla imigrantów z Europy

wschodniej w 2004r.). Ta liberalizacja przepisów imigracyjnych i wydawania pozwoleń na pracę prowadziła do znaczącego (prawie dwukrotnego) zwiększenia się liczby imigrantów. Prasa reprezentująca politykę partii konserwatywnej z kolei podjęła kwestie imigrantów dopiero wówczas, jak można przypuszczać na podstawie list frekwencyjnych i list wystąpień, kiedy stały się one zauważalnym i narastającym problemem społecznym, wpływającym na politykę i gospodarkę kraju. Największa różnica w zwiększeniu zainteresowania sprawami imigrantów zauważalna jest właśnie w przypadku konserwatywnej gazety „The Daily Telegraph”.

Zaskakująca może się wydawać automatyczna ekstrakcja synonimów słowa *immigrant* w systemie SK, którą przedstawia rysunek 1 poniżej. Im większa czcionka tym bardziej bliskoznaczny wyraz z kluczowym *immigrant*. Zatem *immigrant* można zastąpić takimi wyrazami jak dość neutralny *migrant* („migrant”), *refugee* („uchodźca”), *foreigner* („obcokrajowiec”), bardziej nacechowanymi alternatywami, na przykład *businessman*, *worker* („pracownik”) czy *citizen* („obywatel”), ale również raczej kontrowersyjnymi zastępnikami, takimi jak *prisoner* („więzień”), *terrorist* („terrorysta”) czy *criminal* („przestępca”). Naturalnie nie wszystkie wymienione wyrazy można faktycznie uznać za wyrazy bliskoznaczne kluczowego *immigrant*, mimo iż program dostępny na SK nazywa się tezaurs (czyli słownik synonimów). Nawet jednak taka automatyczna ekstrakcja daje pewien obraz tego, w jaki sposób przedstawiony jest imigrant w analizowanej prasie, co może stanowić zaczątek dalszych, bardziej dogłębnych analiz.



Rysunek 1. Automatyczna ekstrakcja synonimów *immigrant*

Dane uzyskane z SK pozwalają zauważyć, w jakich kontekstach używane jest słowo kluczowe. Poniższe zestawienie przedstawia wyniki tej analizy, w których słowo kluczowe (*immigrant/s*) jest w prepozycji do przymiotnika lub w postpozycji w stosunku do frazy nominalnej lub werbalnej. W nawiasach podana jest liczba wystą-

pień w całym korpusie SBP. Wymienione słowa uzyskały najwyższą charakterystykę frekwencyjną w korpusie.

Przymiotnik + „immigrant/s”: illegal (1387), Jewish (118), Polish (84), African (66).

„immigrant/s” + rzeczownik: community (174), worker (102), population (96), labour (49)

„Immigrant/s” + czasownik: come (75), arrive (44), live (42), work (31), settle (20), enter (19), flee (14), die (13), deport (11) etc.

Imigrant kojarzy się zatem Brytyjczykom głównie z nielegalnym pobytem (1387 wystąpień) oraz – w dużej mierze – z Polakami (84 wystąpienia). *Immigrant* współwystępuje z czasownikami i rzeczownikami związanymi z pracą (*labour* – „siła robocza”, *work* – „praca”, *worker* – „pracownik”) oraz z tworzeniem się i zwiększającą się populacją imigrantów (*community* – „społeczność”, *population* – „populacja”, *come* – „przyjeżdżać”, *arrive* – „przyjeżdżać”, *live* – „mieszkać”, *enter* – tu: „wjeżdżać”, *settle* – „osiedlać się”). Można jednak zauważyć, ale w mniejszym stopniu, czasowniki o zabarwieniu negatywnym, np. „uciekać” (*flee*), „umierać” (*die*) i „deportować” (*deport*). I z tych pojęć wyłania się również obraz imigranta. Poniżej przedstawiam kilka przykładów w formie konkordancji w celu przybliżenia nieco szerszego kontekstu.

Konkordancje to kolejne narzędzie oferowane przez językoznawstwo korpusowe. Pozwalają one wyświetlić najbliższy kontekst dla słowa kluczowego, które jest usytuowane w centralnym miejscu w każdym wersie. Najbliższy kontekst widoczny jest natomiast po lewej i prawej stronie od słowa kluczowego i jest automatycznie „przycięty” przez program po konkretnej liczbie znaków. Poniżej przedstawiam kilka wybranych konkordancji.

G,08.10.1993	eme. The French deport certain	immigrants	convicted of criminal activities , but plans
G, 31.05.1993	ion of 32,226. Britain accepted 50,900	immigrants	in the year up to June 1992, 2,300 fewer
DT,11.04.2005	hich makes it difficult to deport illegal	immigrants	or failed asylum seekers . Ask the UN to
DT,05.04.2005	overnment’s hands in deporting illegal	immigrants.	Graham Brady, the shadow Europe
DT,08.02.2005	asionally impossible to deport illegal	immigrants	or failed asylum seekers . –Ask the
DT, 8.07.2005 DT,01.10.2005	SIR – We have 450,000 illegal wers the legal threshold for deporting	immigrants immigrants	on the loose, with our Home Secretary hav holding Dutch residence permits.
G, 01.07.2005	fact that there are	immigrants	living in this country is a

G, 01.03.2005 G,15.07.2005	over 500,000 illegal calculated that there are 500,000 illegal joint charter flights to deport illegal	immigrants immigrants	shocking indictm in Britain. David Trim- ble, the Uster across Europe, the home secretary, Charl
ST,27.02.2005	Daniel McConnell Is deporting illegal	immigrants	from Ireland just a waste of time and m
ST,22.05.2005	territories after threat- ening to deport	immigrants	from the Netherlands Antilles and Arub
G,18.10.2010	restrict Islamic head- gear, and deport	immigrants	with criminal records . The coalition de
G,30.12.2010	ve personally deported 900,000 illegal	immigrants.	Clegg: I win again. Rightwing
DT,18.06.2010	Dutch authorities do not deport illegal	immigrants	from Somalia because the East African cou
T,15.09.2010	antle "illegal camps" and deport illegal	immigrants	was entirely within French and European

Analiza danych w konkordancjach pozwala zauważyć liczby dotyczące imigrantów, o których pisały gazety w różnych okresach. W 1993 r. „The Guardian” pisze o przyjęciu około 50 tysięcy imigrantów (tj. więcej o ok. 20 tysięcy niż liczy mieszkańców miasteczko Lincoln – informacje dostępne po rozszerzeniu lewego kontekstu), czyli o około 2000 mniej niż w poprzednim roku. W 2005 r. „The Daily Telegraph” pisze o 450 tysiącach, a „The Guardian” o 500 tysiącach nielegalnych imigrantów mieszkających w Wielkiej Brytanii. W 2010 r. mowa jest już o 900 tysiącach imigrantów. Liczba imigrantów opisywanych w prasie brytyjskiej ewidentnie zwiększa się zatem na przestrzeni siedemnastu lat.

Kolejna kwestia to występowanie słowa *immigrant* z negatywnymi kolokacjami w postpozycji, takimi jak: „przeszłość kryminalna” (*criminal records*) czy „działania przestępcze” (*criminal activities*). Po głębszej analizie konkordancji z wyrazem *crime* (wyekscerpowanych za pomocą tzw. wyrażen regularnych¹⁰) okazuje się, że kolokacji tych jest jeszcze więcej (150 wystąpień przy ustawieniach do pięciu słów przed słowem kluczowym i 136 wystąpień po słowie kluczowym). Kilka przykładowych konkordancji przedstawiono w aneksie.

Immigrant często jest nośnikiem konotacji związanych z terroryzmem, zwłaszcza w ostatnim czasie (co obrazuje też zaprezentowana wcześniej automatyczna ekstrakcja synonimów). Wpisując stosowne wyrażenia regularne, można wyekstrahować ciąg wyrazów zawierających konkretne słowa, np. *immigrant* występujący w kontekście bardziej ogólnych struktur (ram składniowych) zawierają-

¹⁰ T. Stubblebine, *Regular Expressions*, Pekin 2007.

cych inne słowa, takie jak „terrorysta” lub „terroryzm”¹¹. Użycie wyrażenia regularnego pozwoliło mi uzyskać konteksty, z których kilka prezentuję w aneksie. W całym korpusie SBP znalazłam 49 konkordancji ze słowem kluczowym w kombinacji z wyrazami *terrorist/s* lub *terrorism*. Interesujące jest to, że w tekstach z 1993 r. ta kombinacja wystąpiła tylko trzykrotnie (2 razy w „The Times” i 1 w „The Daily Telegraph”), w 2005 r. aż 34 razy (13 wystąpień w „The Daily Telegraph”, 5 wystąpień w „The Guardian”, 4 wystąpienia w „The Sunday Telegraph”, 2 wystąpienia w „Sunday Times”, 10 wystąpień w „The Times”), a w 2010 r. tylko 12 razy (5 wystąpień w „The Guardian”, 5 wystąpień w „The Daily Telegraph” i 2 wystąpienia w „The Times”). Wyrażenie regularne ze słowem "terroris. *" pozwoliło mi znaleźć w całym korpusie SBP wyrazy *immigrant*, *immigrants* i *immigration* („imigrant”, „imigranci” lub „imigracja”) w kontekście słów *terrorist*, *terrorists* i *terrorism* („terrorysta”, „terroryści” lub „terroryzm”), które występują w bezpośrednim sąsiedztwie lub w sąsiedztwie do pięciu słów w jednej strukturze.

G	attribution to reduction or eradication of	terrorism, illegal immigration	or illegal employment” – nor provide data
ST	assuming two sharing. Concerns over	terrorist risk as passengers miss immigration	check. THE HOME Office is investigating
ST	curity lapse provided an “open door” to	terrorists, immigrants	and asylum seekers. When the 40 passenger
ST	„The flight could have been packed with	terrorists or immigrants	and they would have just been able to walk
ST	ling that this country is an open door to	terrorists, immigrants	and asylum seekers. A spokesman
ST	te co-ordinated UK border force against	terrorism, drug smuggling and immigration	crime. Create a new category of part-time
ST	or looked a bit like terrorists, or knew	terrorists, or was an illegal immigrant	, or was running away and so on.
T	Iraq/the introduction of repressive anti	terrorist legislation/racist immigration	policies. Yes, cynical politicians
T	egative campaigning over issues such as	terrorism, immigration	and crime. Such campaigning could
T	rested show that the problem of Islamic	terrorism is no longer confined to immigrant	communities – seven of those arrested were
T	important issues to deal with, including	terrorism, immigration	, housing, social equality, education, health
T	ent for alleged “links” with international	terrorism as “immigration	detention”. It will be to the Sternal
T	s to arrest people sus-	terrorist acts, tighter	controls and strict bag-

¹¹ [https:// www. sketchengine. co. uk/ user-guide/ user-manual/ concordance-introduction/ regular-expressions/](https://www.sketchengine.co.uk/user-guide/user-manual/concordance-introduction/regular-expressions/)

	pected of planning	immigration	gage inspections
T	were essential if we were going to stop	terrorism, control immigration	and tackle the global criminal conspiracies
T	e need for ID cards as a defence against	terrorism, illegal immigrants	and social security fraud. It was a classic
T	d on criminals, “bogus” asylum-seekers,	terrorists, illegal immigrants	feral children and planning – violating
T	effectively in dealing with issues like	terrorism, crime, immigration	, energy, the war against drugs, or the
T	compulsory scheme as a way of fighting	terrorism, illegal immigration	and identity theft in the aftermath of
T	orders, which helps reduce the threat of	terrorism, crime and immigration	abuse through the electronic collection
T	and oil supplies, a rise in the “level of	terrorists, organized criminals, illegal immigrants	, and illicit goods trying to cross the

Konotacje *immigrant/immigration* z terroryzmem występują zatem najczęściej w korpusie z publikacjami z 2005 r. Liczba wystąpień w tym okresie wynika z pewnością z sytuacji, która zaistniała w Londynie w tamtym czasie, kiedy to nastąpiło kilka zamachów terrorystycznych w różnych częściach stolicy Wielkiej Brytanii. Jeśli chodzi o różnice między poszczególnymi tytułami gazet to zauważalną tendencją jest też większa liczba wystąpień tych kombinacji słów w konserwatywnej “The Daily Telegraph”.

W analizie korpusowej można uzyskać zestawienie kolokatów słowa kluczowego, co w systemie SK można osiągnąć za pomocą narzędzia *Word Sketch*. Rezultaty ekstrakcji danych, w formie skróconej, zamieszczam poniżej (w nawiasie podaję liczbę wystąpień): *Immigrant: illegal* (1387), *family* (113), *worker* (102), *flow* (43), *settle* (20), *amnesty* (15), *terrorist* (13), *deport* (11), *integrate* (10), *hostility* (8) itd.

Nawet takie selektywne przedstawienie kolokatów słowa *immigrant* pozwala zauważyć, że niesie ono konotacje przede wszystkim z nielegalnym procederem przekraczania granicy (*illegal, deport*) i zalewaniem kraju falą imigracji (*flow* – por. konkordancje poniżej), migrowaniem w celu osiedlenia się (często całą rodziną), z polityką państwa goszczącego, której celem jest zintegrowanie imigrantów ze społeczeństwem. Kolejne konotacje dotyczą porażki państw goszczących w ich działaniach integracyjnych (*settle, integrate* – por. konkordancje poniżej), a także przyjazdem w celach zarobkowych lub brakiem woli podjęcia pracy (*work*) przez imigrantów. To kluczowe dla moich badań słowo łączy się również z terroryzmem i wrogością (*hostility*).

G, 5.06.1993	that the government must	immigrants to integrate	into German society and called
DT, 7.04.2005	member of the selected audience who suggested	immigrants would integrate	better if they paid tax and National Insurance
DT, 2.07.2005	What is obnoxious about noticing whether	immigrants or their children integrate	sufficiently to commit themselves to the
DT, 3.10.2005	consolidate its existing bridge	immigrants to integrate	fully. A privileged partnership op
DT, 2.09.2005	monocultural. It was not that the Muslim	immigrant population could not integrate	, but that it chose not to. This is unacceptable
G, 8.11.2005	Equality denies any link between the scale of	immigration and the failure to integrate	but, courageously, has pointed out that
G, 3.09.2005	confront the issue of how	immigrants integrate	into their society. Before she has
T, 28.04.2005	retrenchment has forced painful cuts. Some	immigrants have not only failed to integrate	Into Dutch society, but even actively oppose
T, 3.11.2005	work," he said. "The first generation of	immigrants managed to integrate	better than the third." Both Left
G, 18.10.2010	failure of many of Germany's	immigrants to integrate	was contributing to Germany's
DT, 14.04.2010	Classes' greatest failures. Large groups of	immigrants do not integrate	- and could justifiably argue that they
DT, 30.08.2010	Jews all have the same genes	immigrants cannot integrate	Thilo Sarrazin, a member of the
T, 13.08.2010	citizens. The policy was intended to help	immigrants to integrate	into their local communities. But the exam

Znacząca jest również ekstrakcja kolokatów *flow* („przepływ”, „przyływ”) w konkordancjach z *immigrant(s)/immigration*:

Flowstifle („zdusić”, „zdławić”), *continuing* („ciągły”), *increasing* („wzrastający”), *non-stop*, *relentless* („niepohamowany”); *hinder* („powstrzymać”), *cut* („zmniejszyć”), *halt* („zatrzymać”), *reduce* („zredukować”)

Wybrane przykłady z *flow* w kontekście *immigrant* wyekscerpowane z korpusu przedstawiam poniżej:

- (1) This warning followed a call by John Major at the European summit in Luxembourg for a strong „perimeter fence” around Europe to reduce the **flow of illegal immigrants** [To ostrzeżenie nastąpiło po tym, jak na europejskim szczycie w Luksemburgu John Major nawoływał do stworzenia mocnego „ogrodzenia granicznego” wokół Europy, aby zredukować przepływ nielegalnych imigrantów] („The Guardian”, 17 sierpnia 1993).

- (2) The German authorities believe the slackening of border controls within the EC is increasing the **flow of immigrants** who can move around western Europe more easily in search of work once inside Germany. [Władze niemieckie uważają, że rozluźnienie kontroli granicznej w obrębie UE powoduje wzrost przepływu imigrantów, którzy, będąc już na terenie Niemiec, z większą łatwością mogą się poruszać po zachodniej Europie w poszukiwaniu pracy] („Sunday Times”, 7 marca 1993).

Wspomniany wcześniej kolokator Hask służący do automatycznego wyszukiwania danych w BNC pozwala zauważyć, w jakich kontekstach używane jest słowo *immigrant* w ogólnym języku angielskim. W ten sposób można sprawdzić, czy język gazet (ograniczony w tym wypadku do kilku tytułów i wydań z 3 konkretnych lat) różni się od języka ogólnego. Do analizy posłużył parametr statystyczny LogDice¹² określający stopień prawdopodobieństwa współwystępowania słów kluczowych z ich kolokatami, ten sam bowiem parametr występuje zarówno w systemie SK, jak i Hask. LogDice nie wykazuje także zależności od wielkości korpusu¹³.

Dane uzyskane z wyszukiwarki Hask przedstawia poniższa tabela (ustawienie: do 4 wyrazu w lewym i w prawym kontekście słowa kluczowego):

Tabela 2. Frekwencja wystąpień kolokatów *immigrant/s* i parametr LogDice w korpusie SBT i BNC

SBT		BNC	
<i>Immigrant</i> + przy-miotnik, czasownik, rzeczownik	Frekwencja/ logDice	<i>Immigrant</i> + przy-miotnik, czasownik, rzeczownik	Frekwencja/ logDice
Illegal	1436/11,03	Illegal	98/6,3
Amnesty	74/8	Jewish	31/2,05
Jewish	163/8	Pakistani	5/0,83
Polish	107/7,9	Penniless	4/0,77
Communities	142/7,8	Swamp	5/0,76
Asylum	99/7,8	Craftsman	4/0,75
Influx	56/7,6	Would-be	6/0,72
Would-be	59/7,3	Asian	10/0,72
Second-generation	37/7,2	Soviet	41/0,71
Citizenship	43/7	Irish	18/0,54
Population	125/7	African	11/0,43
Azylum-seekers	33/7	Population	24/0,33
Mexican	45/7	Worker	26/0,29
First-generation	25/6,9	Labour	16/0,23
Flow	63/6,9	Arrive	16/0,22
Jamaican	32/6,8	Settle	7/0,18

¹² <https://www.sketchengine.co.uk/documentation/statistics-used-in-sketch-engine/>

¹³ P. Rychlý, „A lexicographer-friendly association score”, [w:] *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN*, red. naukowa P. Sojka i A. Horák, Brno 2008.

Jak wynika z danych przedstawionych w Tabeli 2. powyżej język gazet brytyjskich różni się od języka ogólnego. Nie są to jednak różnice diametralne. Imigrant kojarzy się przede wszystkim ze słowem *illegal* zarówno w języku ogólnym, jak i w języku omawianych tytułów prasowych. Świadczy o tym wysoka frekwencja występowania tych słów w obu korpusach oraz najwyższy wynik statystyczny LogDice. „Amnestia”, która jest druga na liście frekwencji i parametru LogDice w korpusie gazet, w ogóle nie występuje w BNC. *Jewish* („Żyd”) plasuje się na trzecim miejscu w korpusie gazet i na drugim w korpusie ogólnym, natomiast *would-be* („potencjalny”) jest na ósmym miejscu w korpusie gazet i na siódmym w NKJP. Pozostałe wyrazy w tabeli różnią się charakterystyką frekwencyjną oraz statystyką LogDice. Podobieństwa zatem są widoczne jedynie w zakresie kilku słów. Ponadto, w przypadku języka gazet, korpus SBT zawiera więcej słów oznaczających narodowości imigrantów (w tym Polaków, co wiąże się z wzrostem popularności Wielkiej Brytanii jako kraju emigracji z Polski po 2004 r.). Wysoko plasują się również słowa *communities* („wspólnoty”), *citizenship* („obywatelstwo”) oraz *first-generation* i *second-generation* („pierwszej generacji”, „drugiej generacji”). Kolokaty te podkreślają długość pobytu imigrantów i ich dość silny potencjalny status społeczno-polityczny. Jak wynika z analizy 49 konkordancji jest to status osób ubiegających się o obywatelstwo brytyjskie, a także posiadających już obywatelstwo brytyjskie lub podwójne obywatelstwo. Wymieniony jest też jeden z możliwych powodów ich przybycia, tj. ucieczka przed wojną lub z powodu prześladowań politycznych (*asylum-seekers* – „azylanci”). W BNC z kolei znajdziemy słowa, które wskazują na negatywny odbiór zwiększającego się napływu imigrantów (pejoratywnie nacechowany *swamp* – „zasypywać czymś” vs. bardziej neutralny *influx* – „napływ” występujący w SBT) i niski status materialny imigrantów (*penniless* – „bez grosza”, *labour* – „siła robocza”). Wartość LogDice jest znacząco wyższa w przypadku SBT, co oznacza większe prawdopodobieństwo współwystępowania kolokatów ze słowem kluczowym w języku gazet. Zjawisko to wskazuje na silniejsze konotacje, niż to ma miejsce w korpusie referencyjnym. Prawdopodobieństwo współwystępowania tych słów jest zatem większe w języku gazet niż w języku ogólnym. Informacje te należy traktować jako pewne przybliżenie, gazety w korpusie SBT zawierają bowiem dane z trzech lat (1993, 2005 i 2010), natomiast BNC składa się z tekstów opublikowanych lub zebranych (np. nagrania) do 1995 r. (BNC kompilowany był w latach 1990–1995). Wielka fala imigracji pojawiła się w Wielkiej Brytanii głównie po 2004 r., kiedy kilka państw Europy wschodniej (w tym

Polska) przystąpiło do Unii Europejskiej, co z pewnością spowodowało wzrost liczby publikacji prasowych na temat imigrantów. To zjawisko społeczne nie wystąpiło na taką skalę przed rokiem 1995 (pomijając fale emigracji z byłych kolonii Zjednoczonego Królestwa w latach pięćdziesiątych i sześćdziesiątych), czyli do roku, w którym ukończono tworzenie narodowego korpusu języka angielskiego. Z tego zapewne powodu liczba wystąpień słowa *immigrant* (jak również dlatego, że objętość BNC jest o jedną trzecią mniejsza od korpusu SBT) jest o wiele niższa niż w korpusie SBT. Kolejną kwestią różniącą BNC i SBT jest struktura tych korpusów. BNC zawiera teksty zarówno fikcyjne jak i informacyjne (również inne podtypy tekstów oraz dyskursu, np. język pisany i mówiony). Korpus SBT odnosi się natomiast jedynie do języka pisanego i – zasadniczo – do tekstów informacyjnych. Mimo, iż BNC traktowany jest jako korpus referencyjny to porównanie danych z BNC do danych z korpusu SBT nie może być traktowane jako absolutny wyznacznik badanych parametrów. Z pewnością jednak może on stanowić podstawę do pewnych uogólnień i wskazywania tendencji w analizowanych tekstach.

Ogólne wrażenie, które odnosi się po przejrzeniu list przedstawionych powyżej sugeruje raczej negatywne konotacje związane ze słowem *immigrant* oraz skojarzenie tego pojęcia z aspektem gospodarczym lub społecznym (w odróżnieniu od np. słowa *refugee* („uchodźca”), które występuje w kontekście politycznym). Badanie dyskursu prasowego, który oparty został na danych korpusowych, pozwala zatem określić prozodię semantyczną^{14,15} (tj. tendencję słów do współwystępowania z innymi pod kątem ich parametru aksjologicznego, czyli wartościującego) słów kluczowych poprzez analizę pozytywnego lub negatywnego ładunku emocjonalnego, którego nośnikami są ich kolokaty.

Tabela 3 poniżej ilustruje rozkład najczęściej powtarzających się kolokatów *immigrant(s)* w poszczególnych gazetach. Wyniki te uzyskałam poprzez ekstrakcję danych przeprowadzoną za pomocą *Word Sketch* dla utworzonych specjalnie dla tej analizy podkorpusów SBT („The Times”, „The Guardian” i „The Daily Telegraph”). Z tych danych wynika, że „nielegalny” (*illegal*) i „zabijać” (*kill*) powtarza się najczęściej w „The Guardian”, co może świadczyć o wyrazistym, w dużo większym stopniu negatywnym obrazie imigranta przedsta-

¹⁴ Por. W. Bublitz 1996. *Semantic prosody and cohesive company: somewhat predictable*, „Leuvense Bijdragen, Tijdschrift voor Germaanse Filologie”, 85/1–2: 1–32.

¹⁵ Por. B Lewandowska-Tomaszczyk. 1996. *Cross-linguistic and language-specific aspects of semantic prosody*, „Language Sciences”, 18/1–2: 153–178.

wianym w tej gazecie, niż w innych badanych przeze mnie tytułach prasowych. *Community* („wspólnota”), *group* („grupa”) i *integration* („intergracja”) są również kolokataami najczęściej używanymi w publikacjach „The Guardian”. Zjawisko to z kolei wydaje się wskazywać na określania słowem *immigrant* konkretnej kategorii społecznej (frazą *immigration groups* występuje 17 razy). Rezultaty badań nie zaskakują, gdy uwzględnimy profil pro-społeczny reprezentowany przez „The Guardian”. Pozostałe wyniki analizy są porównywalne.

Tabela 3. Porównanie najczęstszych kolokatów *immigrant* w różnych gazetach

Immigrant	The Guardian	„The Times”	The Daily Telegraph
Illegal	454	399	391
Polish	21	27	27
Number	42	0	56
Employ	18	13	13
Arrest	5	5	5
Deport	4	5	7
Excel	9	3	5
Labour	12	15	9
Worker	24	27	25
Kill	10	0	3
Community	67	38	33
Integration	5	0	0
Flood	5	4	3

Konkluzje

Korpusowa analiza słowa *immigrant(s)* i, w mniejszym stopniu, *immigration* przedstawiona w artykule pozwoliła zauważyć, że badane wyrazy współwystępują w kontekstach głównie nacechowanych pejoratywnie zarówno w korpusie prasy (w korpusie SBT), jak i w korpusie ogólnym (Narodowym Korpusie Brytyjskim). Istotne jest jednak to, że profil imigranta jest nieco inny w tym dotyczącym prasy niż w korpusie referencyjnym. Spośród gazet występujących w badanym materiale SBT, konserwatywna gazeta „The Daily Telegraph” w największym stopniu zwiększyła zainteresowanie problematyką imigracji w latach 2005 i 2010 (w stosunku do 1993 r.). Centro-lewicowa gazeta „The Guardian” na przestrzeni 17 lat podejmowała natomiast problem imigracji w takim samym stopniu, mimo zachodzących w Europie społeczno-politycznych przeobrażeń. „The Guardian” zawiera również najwięcej słów wskazujących na

opis problematyki imigracji w kategoriach nielegalnego procederu przekraczania granicy oraz najczęściej kontekstów z wyrazami opisującymi imigrację jako niezależną grupę i wspólnotę społeczną. Zastosowana w badaniach metodologia analizy korpusowej, dzięki dostępnym narzędziom informatycznym i automatycznej ekstrakcji danych, pozwoliła na ogląd bardzo obszernego materiału prasowego w stosunkowo krótkim czasie i na dostarczenie interesujących danych ilościowych, które stanowią dobry punkt wyjścia do bardziej dogłębnych analiz jakościowych. Analiza korpusowa tekstów prasowych wydaje się być skuteczną i interesującą metodą badawczą w medioznawstwie, która może być wykorzystywana jako alternatywa wobec istniejących już metod badawczych wykorzystywanych w naukach społecznych.

Aneks

G	Le Pen's Basic message – that too many	immigrants cause unemployment and crime	–goes down well in a city with a higher
G	ng the new government to face up to	immigration, crime	Aids and agricultural collapse.
G	rose under a faltering economy, soaring	immigration and a jump in drug-related crime	warders feel a tougher regime will
G	ct criticism of Mr Pasqua who has accused	immigrants of being responsible for the crime	and drug dealing. Christian churches
G	against Mr Pasqua's claims that illegal	immigrants are responsible for crime	drug dealing and has warned his against
G	asked about his allegations that illegal	immigrants were responsible for the crime	wave, Mr Pasqua said the new laws would
G	month, will be used to combat illegal	immigration and crime	Anyone unable to show identity pape
G	outlining his recent initiatives on tougher	immigration controls, fighting crime	and providing greater assistance to the
G	ts campaign on the sensitive issues of	immigration, housing, rising crime	and the economy. The recipe clearly ba
S	rted. Senior police sources working in	immigration-related crime	say that West Africans, particularly Nig
ST	s proposals aimed at controlling illegal	immigration and reducing petty crime	and delinquency. Most controvercially
DT	ordinate its national policies on visas and	immigration, drug control and organized crime	extradiction and the granting of asylum
T	nge of left-wing positions on defence,	immigration, crime	the economy, trade unions and social
T	nt are accentuating antipathy towards	immigrants, a group blamed for crime	And "taking French jobs". The human
T	the main issues in the election were	immigration, crime	and unemployment. On immigration

T	ulate employment and tighten up on	immigration and crime	It is likely to clamp down on “family
T	ich also promised tougher enforcement of	immigration laws and a campaign against crime	focused on the government’s plan for pulling
T	ight protection, repatriation of illegal	immigrants and crime	vention. Three other agreements were
D	ile the Conservatists have led on both	immigration and crime	However, as the figures show,
D	and again – during the Iraq war, over	immigration, violent crime	And hospital waiting-list figures. It has
G	erate tone of the policy document on	crime and immigration	was the clearest indication that the ort
G	g for the prosecution of rape as a war	crime, for the lifting of immigration	restrictions on women war refugees an
G	jobs, defending farmers against Gatt	crime, immigration	and the family. In the first round
G	ging human rights in the fight against	crime and illegal immigration	Mr Paqua’s regret that police powers
G	ok controversial stands on civil rights,	crime and immigration	few have broken through into comm
G	to announce tougher measures to combat	crime and steps to integrate immigrants	when he makes a statement to parliament
G	evious arguments have urgently cited	crime prevention and immigration	control. They also favoured voluntary n
G	s powerfully and mendaciously linked	crime and immigration	it is a big winner and loser of votes.
G	is of consent. Tax, welfare, inequality;	crime, immigration	drugs, schooling: each of these is the
D	an invaluable aide in the fight against	crime, illegal immigration	and under-age drinking. But it would b
D	r elements of the agreement include	crime, repatriating immigrants	and fishery disputes. Israeli tanks
D	creasingly sullen about such things as	crime and immigration	facets of a spreading anxiety about th
D	other uses, including the fight against	crime as well as immigration	and even taxation matters. Mr John
T	police. Faced with an alarming rise in	crime, especially among immigrants	and asylum seekers, Rudolf Seiters, the
T	terday and announce a crackdown on	crime and illegal immigration	His style, depicted as an “iron fist”
S	e’s obsession with a different kind of	crime: illegal immigration	To the dismay of the French governm
S	ake the fight against drugs, organized	crime and illegal immigration	on to the same international level whe
T	form and the passing of tough bills on	crime and immigration	will need cross-party consensus, as lib
T	nder of his rival, Michael Howard, on	crime and immigration	In a phrase that should please the mo

Bibliografia

- Baker P. (2006), *Using Corpora in Discourse Analysis*. Londyn.
- Bączkowska A. (2003), *Wybrane techniki wykorzystania korpusów językowych w nauczaniu języka angielskiego*, *Języki Obce w Szkole* 3, s. 64–76.
- Bublitz W. (1996), *Semantic prosody and cohesive company: somewhat predictable*, "Leuvense Bijdragen, Tijdschrift voor Germanse Filologie", 85/1–2: 1–32.
- Lewandowska-Tomaszczyk B. (1996), *Cross-linguistic and language-specific aspects of semantic prosody*, "Language Sciences", 18/1–2: 153–178.
- Lewandowska-Tomaszczyk B. (2005), *Podstawy językoznawstwa korpusowego*. Łódź.
- Narodowy Korpus Języka Polskiego* (2012), A. Przepiórkowski, M. Bańko, R. Górski, B. Lewandowska-Tomaszczyk (red.), PWN, Warszawa.
- Pęzik P. (2013), *Paradygmat dystrybucyjny w badaniach frazeologicznych. Powtarzalność, reprodukcja, idiomatyzacja*, [w:] *Metodologie językoznawstwa. Ewolucja języka. Ewolucja teorii językoznawczych*, red. naukowa P. Stalmaszczyk, Łódź.
- Rychly P. (2008), *A lexicographer-friendly association score*, [w:] *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN*, red. nauk. P. Sojka i A. Horák, Brno.
- Stubblebine T. (2007), *Regular Expressions*. Pekin.
- Tognini-Bonelli T. (2001), *Corpus Linguistics at Work*, John Benjamins, Amsterdam.

Źródła internetowe

- [https:// www. sketchengine. co. uk/ user-guide/ user-manual/ concordance-intro-duction/ regular-expressions/](https://www.sketchengine.co.uk/user-guide/user-manual/concordance-introduction/regular-expressions/)
- <https://www.sketchengine.co.uk/documentation/statistics-used-in-sketch-engine/>
- pelcra.pl/hask_pl/home
- the.sketchengine.co.uk