

Monika Szela

Wyższa Szkoła Filologiczna we Wrocławiu

O wykorzystaniu Angielsko-Polskiego Korpusu Równoległego Tekstów Prawnych w badaniu cech języka tekstów tłumaczonych

On using a English-Polish Parallel Corpus of Legal Texts in research on features of the translational language

Abstract

This paper aims to present the compilation of the English-Polish parallel corpus comprising legal acts of the European Union as well as preliminary investigations into this corpus against comparative corpora of British and Polish acts. The corpora have been compiled for the purposes of research on grammatical and lexical features of translated texts in comparison to texts produced originally by native speakers of the target language. The phenomenon of divergence between translated and non-translated texts have recently been given considerable attention and this research is intended to contribute to the studies dedicated to this phenomenon. Therefore, both translated and non-translated legal acts have been acquired.

The parallel corpus contains two subcorpora with English and Polish texts of the EU body of law (L series) published by the European Commission from 2004 to 2011. The size of each subcorpus exceeds forty million words. The texts were downloaded both as plain text files and aligned translation memories. Additionally, two comparative corpora, covering the same period of time, were compiled: the first being the general legal acts of the British Parliament, and the second being legal acts published in the Polish Journal of Laws. All the files underwent basic – though labour-intensive – processing: pdf files were converted to plain text formats, and character encoding was unified if required. The files were then uploaded to WordSmith Tools, a tool for text analysis, which produced word frequency lists and key word lists.

The initial analyses included the investigation of (i) a hand-picked Polish improper verb *należy* with an untypical frequency in the parallel corpus against the distribution of the English modal verb *shall* in both parallel and comparative corpora; and (ii) Polish impersonal verb forms ended with *-no*, *-to*. As far as the impersonals were concerned, it was assumed that the analysis would confirm the

under-representation of these forms in translated texts; however, the results reveal no such tendency. The paper ends with tentative conclusions drawn from the results as more detailed study into thus compiled corpora is called for.

Słowa kluczowe: korpus równoległy, korpus porównawczy, język polski, język angielski, język przekładu, interferencja, uniwersalia tłumaczeniowe, język prawniczy

Keywords: parallel corpus, comparative corpus, Polish, English, translationese, interference, translation universals, legalese

1. Wprowadzenie

Omówione w niniejszym artykule zadanie utworzenia równoległego korpusu angielsko-polskiego tekstów prawych jest elementem badań nad zjawiskiem nienaturalności języka tekstu przekładu i jego cech na materiale aktów prawnych. Zagadnieniu braku naturalności tekstów tłumaczonych w porównaniu z tekstami tworzonymi oryginalnie przez rodzimych użytkowników języka docelowego (Olohan, 2004: 90) poświęcono już wiele badań, a w literaturze przedmiotu pojawiły się nowe terminy, jak język tłumaczeń, język hybrydowy czy trzeci kod (ang. translationese, translatese, hybrid language, third code). Wśród przyczyn wspomnianego zjawiska badacze podkreślają zarówno wpływ języka źródłowego, jak i uniwersalne tendencje niezależne od pary językowej, inherentne dla samego procesu tłumaczenia, mające wpływ na produkt tłumaczenia (Mauranen, 2008: 33-34). Od lat dziewięćdziesiątych ubiegłego wieku podejmowano badania dotyczące nie tylko interferencji, ale stawiano również hipotezy o istnieniu uniwersaliów tłumaczeniowych wpisanych w proces tłumaczenia. Pojawiły się między innymi hipotezy o eksplicytacji w przekładzie, uproszczeniu składniowym, konwencjonalizacji języka przekładu, niedoprezentacji struktur języka docelowego w przypadku ich niewystępowania w języku źródłowym czy konwergencji. Badacze są zgodni, że brak naturalności produktu tłumaczenia w porównaniu z tekstami oryginalnie utworzonymi w języku docelowym nie sprowadza się tylko do niekompetencji językowej tłumacza czy niewystarczającej wiedzy przedmiotowej, ponieważ pojawia się również w tłumaczeniach wykonanych przez profesjonalnych tłumaczy.

W celu określenia cech języka przekładu i rozróżnienia, które z nich są rezultatem interferencji, a które należą do tendencji uniwersalnych dla procesu tłumaczenia, wykorzystuje się metody językoznawstwa korpusowego, które są użytecznym narzędziem badania rzeczywistego uzusu języka i obserwacji zachodzących regularności dla dużych ilości tekstów (Piotrowski, 2003: 143-147). Dzięki korpusom zyskuje się dane frekwencyjne dotyczące dystrybucji poszczególnych słów i wyrażeń w tekstach tłumaczonych oraz typowych

połączeń wyrazowych. Takie dane zestawia się z odpowiednimi danymi dla tekstów nietłumaczonych, powstających w języku docelowym.

Przekładoznawstwo opisowe posługuje się korpusami równoległymi, które zawierają tekst źródłowy zestawiony z jego tłumaczeniem (tłumaczeniami), korpusami porównawczymi, które zawierają teksty o podobnym profilu (na przykład prawne, medyczne) w różnych językach (Mauranen, 2008: 33), a także korpusami referencyjnymi, czyli zbiorami tekstów nietłumaczonych, które służą za odniesienie przy badaniu cech tekstu tłumaczonego. Korpus równoległy pozwala na badanie tłumaczeń pod kątem ewentualnej interferencji, zaś korpus porównawczy umożliwia badanie tendencji uniwersalnych oraz cech, które sprawiają, że tekst nie brzmi naturalnie w porównaniu z tekstami oryginalnymi. Na potrzeby analizy produktu tłumaczenia ważne okazuje się nie tylko porównanie tekstu źródłowego z tekstem docelowym, ale również porównanie tekstu docelowego z korpusem tekstów utworzonych oryginalnie w języku docelowym (por. Grabowski, 2011: 89-112).

Na potrzeby niniejszego badania, które dotyczy cech języka tłumaczeń utworzono zarówno korpusy równoległe zawierające tłumaczenia tekstów prawnych na język polski wraz z ich tekstami źródłowymi w języku angielskim, jak i korpusy porównawcze tekstów prawnych tworzonych oryginalnie w języku polskim i angielskim. W artykule przedstawione zostaną pokrótce wymienione korpusy, proces ich kompilacji oraz wstępne analizy. Zamierzeniem autorki jest kontynuacja badań na przedstawionym materiale, których celem jest weryfikacja hipotezy o odmienności gramatycznej języka tekstów tłumaczonych w porównaniu z tekstami nietłumaczonymi. Omawiane badania będą mogły stanowić przyczynek do badań aktualnie prowadzonych dla podobnych zbiorów tekstów (Biel, 2014).

2. Opis korpusów

Materiał do badań ograniczono do jednego rodzaju tekstów aktów prawnych, a mianowicie ustaw i rozporządzeń. Na potrzeby analizy cech języka tłumaczeniowego, skompilowane zostały następujące korpusy tekstów prawnych:

- korpus równoległy aktów prawnych Unii Europejskiej tłumaczonych z języka angielskiego na język polski: teksty źródłowe w języku angielskim (dalej KRAN) oraz odpowiadające im teksty docelowe w języku polskim (dalej KRPL);
- dwa korpusy porównawcze: teksty ustaw i rozporządzeń oryginalnie napisane w języku polskim (KPPL) oraz teksty brytyjskich aktów prawnych oryginalnie napisane w języku angielskim (KPAN).

W tabeli 1. przedstawiono podstawowe statystyki dla wykorzystanych korpusów równoległych i porównawczych (liczba okazów oznacza liczbę wszystkich wyrazów w korpusie, zaś liczba typów oznacza liczbę różnych wyrazów w korpusie).

Korpus	Liczba plików	Liczba okazów	Liczba typów
KRAN	76 851	49 745 496	191 939
KRPL	75 820	40 200 352	318 644
KPAN	357	12 947 371	18 906
KPPL	774	3 266 783	53 890

Tabela 1. Podstawowe dane korpusu równoległego (KRAN – KRPL) oraz korpusów porównawczych (KPAN oraz KPPL)

2.1 Korpus równoległy. Informacje szczegółowe

Korpus równoległy stanowią zbiory tekstów aktów prawnych¹ publikowanych przez Generalny Dyrektorat ds. Tłumaczeń Unii Europejskiej. Akty prawne UE zostały wybrane z kilku powodów: po pierwsze, ze względu na ogromną bazę tekstów udostępnioną publicznie, a po drugie ze względu na brak odmienności systemów prawnych (szczególnie w okresie poakcesyjnym, gdy niektóre z aktów prawnych, np. rozporządzenia, mają tę samą moc prawną w kraju języka docelowego). Trzecim powodem jest wysoka jakość tłumaczeń wykonywanych w ramach konsultacji językowych między poszczególnymi komisjami, co eliminuje problem kompetencji tłumacza (a raczej ewentualnego jej braku).

Podkorpus angielski zawiera akty z Dziennika Urzędowego serii L z lat 2004 – 2011 pobrane z oficjalnych stron UE² (EUR-Lex): korpus równoległy DGT-Acquis³. Pliki w formacie tekstowym⁴ zostały pobrane w postaci skompresowanych plików, odrębnie dla poszczególnych lat. Łącznie pobrano osiem plików, z których po rozpakowaniu otrzymano 76 851 plików tekstowych gotowych do badania za pomocą specjalnego oprogramowania do badania korpusów tekstów WordSmith Tools 6.0⁵. Podkorpus angielski jest traktowany jako zbiór tekstów źródłowych tłumaczenia⁶.

1 http://ec.europa.eu/legislation/index_en.htm (ED: 22/01/2015)

2 <http://eur-lex.europa.eu/homepage.html>

3 <http://ipsc.jrc.ec.europa.eu/index.php?id=783>

4 <http://ec.europa.eu/jrc/en/language-technology/dgt-acquis/da1-ft>

5 <http://www.lexically.net/wordsmith>

6 Jest tak mimo zastrzeżeń wynikających z wieloetapowego procesu sporządzania dokumentów. Na początku tego procesu tworzona jest wersja robocza przekazywana do komisji językowych, które po zapoznaniu się zgłaszają problemy związane z tłumaczeniem, przekładające się dalej na zmiany w drugiej wersji roboczej, ponownie przekazywanej do komisji językowych (Steinberger et al., 2013: 455). Ten tryb tworzenia aktów prawnych powoduje, że obecnie nie uznaje się żadnego języka ani za język źródłowy, ani za język docelowy (Biel, 2014: 5)

Zbiorem tekstów tłumaczonych na język polski jest podkorpus aktów prawnych tworzących serię L Dziennika Urzędowego w języku polskim. Teksty zostały pobrane ze stron zawierających oficjalne publikacje EU z lat 2004 – 2011. Pliki pobrano również w formacie tekstowym⁷ w postaci ośmiu skompresowanych plików. Po rozpakowaniu otrzymano 75 820 plików tekstowych. Ponieważ jednak kodowanie tekstów UTF-8 powodowało błędy podczas badania za pomocą WordSmith Tools, pliki zostały przekonwertowane za pomocą komercyjnej aplikacji UTF Cast Professional⁸ na UTF-16LE. Po przekonwertowaniu, teksty można było analizować za pomocą WordSmith Tools.

Na potrzeby badania interferencji językowej pobrano również korpus wyrównany JRC-Acquis⁹ oraz pamięci tłumaczeniowe DGT-TM¹⁰.

2.2 Korpusy porównawcze. Informacje szczegółowe

2.2.1 Korpus brytyjskich aktów prawnych

Na korpus tekstów brytyjskich składają się ustawy Parlamentu Brytyjskiego (Public General Acts) z lat 2004-2014 publikowane w czasie zbliżonym do okresu publikacji dla korpusu równoległego UE. Teksty pochodzą z oficjalnej strony zawierającej brytyjskie akty prawne¹¹. Pobranym zostało 357 plików w formacie pdf, które następnie zostały przekonwertowane do formatu tekstowego za pomocą wbudowanej aplikacji czytelnika Adobe Reader 9.

2.2.2 Korpus polskich tekstów prawnych

Teksty nietłumaczone zawierające akty prawne napisane oryginalnie przez polskich prawników były pobrane z oficjalnych stron zawierających polskie ustawy: Dziennik Ustaw¹² oraz Internetowego Systemu Aktów Prawnych¹³. Teksty były filtrowane według lat oraz rodzaju: pobierane były ustawy z lat 2004-2011 w formacie pdf. Po skomplikowanej procedurze konwertowania plików do formatu odczytywanego przez WordSmith Tool, pozyskano 774 pliki tekstowe.

7 <http://ec.europa/jrc/en/language-technology/dgt-acquis/da1-ft>

8 <http://www.rotatingscrew.com/utfcast-professional.aspx>

9 <http://optima.jrc.it/Acquis/JRC-Acquis.3.0/corpus/>

10 <http://open-data.europa.eu/en/data/dataset/dgt-translation-memory>

11 <http://www.legislation.gov.uk/>

12 <http://dziennikustaw.gov.pl/>

13 <http://isap.sejm.gov.pl/index.jsp>

2.3 Korpusy referencyjne

W docelowym badaniu wykorzystane zostaną również dwa korpusy referencyjne: Narodowy Korpus Języka Polskiego¹⁴ (NKJP) oraz Narodowy Korpus Języka Brytyjskiego (British National Corpus, BNC) pobrany w całości po otrzymaniu zgody. Według statystyk przedstawionych na stronie BNC, korpus zawiera sto milionów słów¹⁵.

3. Badania wstępne

Skompilowane korpusy stanowią podstawowy materiał do badania cech języka przekładu. Głównym korpusem podlegającym badaniu jest podkorpus tekstów prawnych tłumaczonych na język polski. Dla tego podkorpusu są analizowane cechy języka tłumaczeń pod kątem ich odmienności od tekstów nietłumaczonych. Pozostałe korpusy są korpusami pomocniczymi pozwalającymi na określenie zakresu ewentualnej interferencji z języka źródłowego oraz tendencji niezależnych od wpływu tekstów źródłowych, charakterystycznych dla procesu tłumaczenia.

3.1. Listy frekwencyjne

Pierwszym elementem badań jest utworzenie list frekwencyjnych oraz list słów kluczowych dla korpusów równoległych i ich dystrybucja w korpusach porównawczych za pomocą programu WordSmith Tools 6.0. Lista frekwencyjna zawiera listę słów uporządkowanych według liczby wystąpień w korpusie. Utworzono cztery listy frekwencyjne: dwie dla korpusu równoległego oraz dwie listy dla korpusów porównawczych. W tabelach 2.-5. przedstawiono listy frekwencyjne z wynikami dla pierwszych stu najczęstszych słów dla każdego korpusu. Czcionką pogrubioną zaznaczono elementy analizowane w dalszej części artykułu (ponieważ korpusy zostały poddane tylko częściowej obróbce, na listach mogą pojawiać się wyrazy z błędnym zapisem lub kodowaniem; # zastępuje cyfry pojawiające się w korpusach).

Lp.	Słowo	Liczba w KRPL	% w KRPL	Lp.	Słowo	Liczba w KRPL	% w KRPL
1	#	3197162	7,19	51	NIŻ	66045	0,15
2	W	1814379	4,08	52	ICH	65455	0,15
3	I	977858	2,20	53	TO	64532	0,15
4	Z	879833	1,98	54	EUROPEJSKIEJ	64314	0,14

14 <http://nkjp.pl/>

15 <http://www.natcorp.ox.ac.uk/>

Lp.	Słowo	Liczba w KRPL	% w KRPL	Lp.	Słowo	Liczba w KRPL	% w KRPL
5	DO	684939	1,54	55	ŚRODKÓW	60396	0,14
6	NA	608112	1,37	56	UWZGLĘDNIAJĄC	60291	0,14
7	M	552688	1,24	57	CZŁONKOWSKIE	59592	0,13
8	Ý	540161	1,21	58	SZCZEGÓLNOŚCI	59412	0,13
9	LUB	370437	0,83	59	UNII	59307	0,13
10	R	303742	0,68	60	TE	58578	0,13
11	WE	301690	0,68	61	MOWA	58494	0,13
12	SIĘ	299330	0,67	62	PRODUKTÓW	57623	0,13
13	O	284655	0,64	63	DOTYCZĄCE	56390	0,13
14	NR	241323	0,54	64	RAMACH	54346	0,12
15	ORAZ	239200	0,54	65	JEŻELI	53734	0,12
16	ART.	237609	0,53	66	ZAŁĄCZNIKU	52989	0,12
17	NIE	228996	0,52	67	POMOCY	52972	0,12
18	DNIA	224912	0,51	68	JAKO	52006	0,12
19	PRZEZ	200415	0,45	69	JAK	50987	0,11
20	JEST	187938	0,42	70	TEGO	50840	0,11
21	DLA	182632	0,41	71	NINIEJSZEGO	50672	0,11
22	A	169878	0,38	72	MOGĄ	50659	0,11
23	UST	160874	0,36	73	EWG	50001	0,11
24	ŻE	154807	0,35	74	DECYZJI	49621	0,11
25	ZGODNIE	149402	0,34	75	L	48663	0,11
26	ROZPORZĄDZENIA	145945	0,33	76	TAKŻE	47884	0,11
27	KOMISJI	136219	0,31	77	U	47689	0,11
28	SĄ	134635	0,30	78	ZAKRESIE	47547	0,11
29	OD	132944	0,30	79	INNYCH	47089	0,11
30	ROZPORZĄDZENIE	127015	0,29	80	LIT	46453	0,10
31	ARTYKUŁ	123113	0,28	81	CZŁONKOWSKICH	46032	0,10
32	RADY	109102	0,25	82	TYCH	45938	0,10
33	ZA	92306	0,21	83	PRZY	45017	0,10
34	SPRAWIE	91447	0,21	84	PAŃSTW	45005	0,10
35	KTÓRE	91382	0,21	85	WSPÓLNOTY	45002	0,10
36	CELU	88097	0,20	86	B	44411	0,10
37	NALEŻY	87873	0,20	87	NINIEJSZE	44333	0,10
38	PAŃSTWA	87829	0,20	88	DYREKTYWY	43785	0,10
39	PRZYPADKU	87294	0,20	89	POD	43422	0,10
40	BYĆ	85994	0,19	90	MIĘDZY	43304	0,10
41	KTÓRYCH	84366	0,19	91	DECYZJA	42245	0,10
42	TYM	78470	0,18	92	RYNKU	42061	0,09
43	PO	75598	0,17	93	C	41442	0,09
44	ŚRODKI	74514	0,17	94	WARTOŚCI	41310	0,09
45	JEGO	71196	0,16	95	RÓWNIEŻ	40922	0,09
46	ZE	71124	0,16	96	EUROPEJSKIEGO	40674	0,09

Lp.	Słowo	Liczba w KRPL	% w KRPL	Lp.	Słowo	Liczba w KRPL	% w KRPL
47	ODNIESIENIU	70692	0,16	97	ZAŁĄCZNIK	39995	0,09
48	KOMISJA	70687	0,16	98	II	39842	0,09
49	MOŻE	67202	0,15	99	WSZYSTKICH	39775	0,09
50	CO	67191	0,15	100	PKT	39753	0,09

Tabela 2. Lista frekwencyjna dla podkorpusu aktów prawnych UE tłumaczonych na język polski (KRPL)

Lp.	Słowo	Liczba w KRAN	% w KRAN	Lp.	Słowo	Liczba w KRAN	% w KRAN
1	THE	4105230	8,25	51	ARTICLEÂ	88134	0,18
2	#	3233168	6,50	52	STATE	86267	0,17
3	OF	2495090	5,01	53	FOLLOWING	85843	0,17
4	TO	1284276	2,58	54	INFORMATION	85736	0,17
5	AND	1272934	2,56	55	ALL	84191	0,17
6	IN	1238765	2,49	56	HAVE	81288	0,16
7	FOR	691226	1,39	57	MEASURES	78550	0,16
8	A	576733	1,16	58	REFERRED	77864	0,16
9	BE	506839	1,02	59	DIRECTIVE	77412	0,16
10	OR	426358	0,86	60	THAN	76652	0,15
11	ON	423319	0,85	61	INTO	76428	0,15
12	SHALL	389849	0,78	62	OUT	76189	0,15
13	BY	384255	0,77	63	AGREEMENT	75886	0,15
14	WITH	326489	0,66	64	I	74921	0,15
15	IS	326365	0,66	65	IF	74861	0,15
16	REGULATION	323248	0,65	66	UNION	74057	0,15
17	THAT	318328	0,64	67	SUCH	73522	0,15
18	AS	297674	0,60	68	REGARD	72444	0,15
19	THIS	292751	0,59	69	HAVING	72047	0,14
20	ARTICLE	265127	0,53	70	THEIR	71773	0,14
21	EC	264468	0,53	71	FINANCIAL	71178	0,14
22	COMMISSION	225198	0,45	72	NOÂ	70215	0,14
23	Â€	220035	0,44	73	PARTICULAR	68103	0,14
24	NO	206352	0,41	74	WHERE	67969	0,14
25	FROM	197610	0,40	75	PROVIDED	67894	0,14
26	EUROPEAN	197566	0,40	76	Ë	67091	0,13
27	NOT	197241	0,40	77	WAS	63104	0,13
28	WHICH	175284	0,35	78	BEEN	62998	0,13
29	ARE	169095	0,34	79	MARKET	62329	0,13
30	MEMBER	160637	0,32	80	BETWEEN	61611	0,12
31	AT	154803	0,31	81	AID	60085	0,12
32	Î	154013	0,31	82	WITHIN	59378	0,12
33	AN	149751	0,30	83	DOWN	56087	0,11

Lp. Słowo	Liczba w KRAN	% w KRAN	Lp. Słowo	Liczba w KRAN	% w KRAN
34 COMMUNITY	138744	0,28	84 PERIOD	53885	0,11
35 IT	136235	0,27	85 P	53789	0,11
36 OTHER	135076	0,27	86 SET	52640	0,11
37 COUNCIL	130437	0,26	87 THOSE	51529	0,10
38 MAY	120278	0,24	88 EEC	51272	0,10
39 STATES	116105	0,23	89 PART	50866	0,10
40 DECISION	114810	0,23	90 USED	49706	0,10
41 ITS	111997	0,23	91 MUST	49646	0,10
42 ANNEX	110035	0,22	92 CERTAIN	49546	0,10
43 UNDER	100926	0,20	93 CONDITIONS	49229	0,10
44 Ę	96245	0,19	94 ONE	48976	0,10
45 SHOULD	94760	0,19	95 APPLICABLE	48829	0,10
46 ACCORDANCE	94480	0,19	96 USE	48010	0,10
47 	93647	0,19	97 THEY	46835	0,09
48 HAS	91559	0,18	98 TREATY	46787	0,09
49 ANY	91557	0,18	99 RULES	46747	0,09
50 PRODUCTS	89761	0,18	100 PROVISIONS	46476	0,09

Tabela 3. Lista frekwencyjna dla podkorpusu aktów prawnych UE w języku angielskim (KRAN)

Lp. Słowo	Liczba w KPPL	% w KPPL	Lp. Słowo	Liczba w KPPL	% w KPPL
1 #	549333	16,82	51 POLSKIEJ	4269	0,13
2 W	143867	4,40	52 STOSUJE	4205	0,13
3 Z	69089	2,11	53 USTAWIE	4161	0,13
4 I	68766	2,11	54 WŁAŚCIWY	4050	0,12
5 NR	59366	1,82	55 DZIAŁ	4019	0,12
6 POZ	57620	1,76	56 DZIAŁALNOŚCI	3985	0,12
7 ART	56017	1,71	57 PAŃSTWA	3968	0,12
8 O	54992	1,68	58 ZE	3864	0,12
9 UST	42121	1,29	59 TERMINIE	3824	0,12
10 DO	41320	1,26	60 RZECZYPOSPOLITEJ	3799	0,12
11 NA	37690	1,15	61 MINISTER	3775	0,12
12 R	35369	1,08	62 JEGO	3752	0,11
13 LUB	31131	0,95	63 INFORMACJI	3732	0,11
14 SIĘ	26904	0,82	64 OSÓB	3701	0,11
15 ORAZ	25062	0,77	65 PRZEPISÓW	3658	0,11
16 MOWA	24230	0,74	66 ZOSTAŁY	3606	0,11
17 DNIA	18440	0,56	67 ŻYCIE	3583	0,11
18 PKT	16521	0,51	68 DRODZE	3576	0,11
19 PRZEZ	14686	0,45	69 S	3564	0,11
20 USTAWY	14607	0,45	70 WYMIENIONEJ	3560	0,11

Lp. Słowo	Liczba w KPPL	% w KPPL	Lp. Słowo	Liczba w KPPL	% w KPPL
21 KTÓRYCH	12884	0,39	71 ZM	3555	0,11
22 NIE	12859	0,39	72 PRACY	3550	0,11
23 A	11680	0,36	73 PÓŻN	3529	0,11
24 BRZMIENIE	11439	0,35	74 PRZEPISY	3432	0,11
25 OTRZYMUJE	10367	0,32	75 OCHRONY	3416	0,10
26 OD	10200	0,31	76 ŚRODKÓW	3339	0,10
27 KTÓRYM	10183	0,31	77 ROZPORZĄDZENIA	3300	0,10
28 DZ	9758	0,30	78 BYĆ	3279	0,10
29 PO	8979	0,27	79 PRZED	3275	0,10
30 U	8842	0,27	80 OSOBY	3250	0,10
31 JEST	8647	0,26	81 TYCH	3192	0,10
32 ZA	7338	0,22	82 OKREŚLONYCH	3161	0,10
33 ALBO	6976	0,21	83 DANYCH	3132	0,10
34 KTÓREJ	6840	0,21	84 WNIOSEK	3113	0,10
35 MOŻE	6728	0,21	85 TAKŻE	3092	0,09
36 PRZYPADKU	6572	0,20	86 DECYZJI	3058	0,09
37 BRZMIENIU	6307	0,19	87 KONTROLI	3013	0,09
38 SPRAW	6278	0,19	88 C	3006	0,09
39 JEŻELI	6182	0,19	89 UMOWY	2968	0,09
40 B	6133	0,19	90 DNI	2967	0,09
41 ZAKRESIE	6087	0,19	91 POSTĘPOWANIA	2964	0,09
42 DLA	5812	0,18	92 PRAWO	2960	0,09
43 PODSTAWIE	5718	0,18	93 OGŁOSZONE	2955	0,09
44 DODAJE	5354	0,16	94 SPOSÓB	2947	0,09
45 ZMIANY	5334	0,16	95 ODPOWIEDNIO	2943	0,09
46 USTAW	5329	0,16	96 PRAWA	2943	0,09
47 DZIENNIK	4951	0,15	97 RADY	2937	0,09
48 NIŻ	4781	0,15	98 WE	2927	0,09
49 TYM	4700	0,14	99 TEGO	2864	0,09
50 ICH	4285	0,13	100 CZĘŚĆ	2842	0,09
			275 NALEŻY	1054	0,03

Tabela 4. Lista frekwencyjna dla korpusu porównawczego polskich aktów prawnych

Lp. Słowo	Liczba w KPAN	% w KPAN	Lp. Słowo	Liczba w KPAN	% w KPAN
1 #	1184510	9,13	51 SUCH	32630	0,25
2 THE	885054	6,82	52 MUST	32451	0,25
3 OF	619758	4,78	53 FROM	28961	0,22
4 A	417996	3,22	54 AT	28616	0,22
5 TO	396460	3,05	55 STATE	28358	0,22
6 IN	373462	2,88	56 AMOUNT	25802	0,20

Lp.	Słowo	Liczba w KPAN	% w KPAN	Lp.	Słowo	Liczba w KPAN	% w KPAN
7	AND	252007	1,94	57	APPLIES	25701	0,20
8	OR	246292	1,90	58	INSERT	25569	0,20
9	SECTION	191003	1,47	59	SECRETARY	25229	0,19
10	FOR	188061	1,45	60	WHERE	24514	0,19
11	IS	170875	1,32	61	REGULATIONS	24160	0,19
12	ACT	148456	1,14	62	MEANS	23785	0,18
13	B	126381	0,97	63	BEFORE	23677	0,18
14	BY	124733	0,96	64	SUB	23228	0,18
15	C	118370	0,91	65	RESPECT	22750	0,18
16	THAT	113097	0,87	66	HAVE	21956	0,17
17	AS	112392	0,87	67	WITHIN	21703	0,17
18	UNDER	111840	0,86	68	MAKE	21678	0,17
19	BE	99333	0,77	69	INCOME	21628	0,17
20	ANY	90649	0,70	70	CHAPTER	21141	0,16
21	AN	89753	0,69	71	COURT	20221	0,16
22	SUBSECTION	89306	0,69	72	CASE	19818	0,15
23	THIS	88276	0,68	73	TIME	19537	0,15
24	PARAGRAPH	83076	0,64	74	NOTICE	19434	0,15
25	WHICH	75625	0,58	75	EFFECT	19388	0,15
26	ON	69176	0,53	76	LOCAL	19357	0,15
27	PERSON	68520	0,53	77	D	19343	0,15
28	MAY	67025	0,52	78	SPECIFIED	19176	0,15
29	WITH	61317	0,47	79	SCHEME	19111	0,15
30	PART	60765	0,47	80	BEEN	18989	0,15
31	NOT	58924	0,45	81	SERVICES	18324	0,14
32	IF	57971	0,45	82	SO	18323	0,14
33	IT	47659	0,37	83	INFORMATION	16870	0,13
34	SCHEDULE	43991	0,34	84	MEANING	16588	0,13
35	MADE	42796	0,33	85	I	16552	0,13
36	ORDER	40841	0,31	86	WHO	16360	0,13
37	RELATION	40308	0,31	87	REFERENCE	16003	0,12
38	HAS	39570	0,30	88	GIVEN	15635	0,12
39	ARE	39529	0,30	89	APPLICATION	15591	0,12
40	AFTER	38494	0,30	90	PERSONS	15322	0,12
41	OTHER	38090	0,29	91	APPLY	15292	0,12
42	PROVISION	37870	0,29	92	SUBJECT	15194	0,12
43	PURPOSES	37237	0,29	93	HEALTH	15030	0,12
44	AUTHORITY	35531	0,27	94	POWER	15006	0,12
45	S	35095	0,27	95	SERVICE	14873	0,11
46	PERIOD	34562	0,27	96	YEAR	14743	0,11
47	TAX	34457	0,27	97	FUNCTIONS	14461	0,11
48	COMPANY	33324	0,26	98	THAN	14329	0,11

Lp. Słowo	Liczba w KPAN	% w KPAN	Lp. Słowo	Liczba w KPAN	% w KPAN
49 RELEVANT	33090	0,25	99 OFFENCE	13933	0,11
50 SUBSTITUTE	32860	0,25	100 ITS	13888	0,11
			103 SHALL	13279	0,10
			737 SHOULD	1763	0,01

Tabela 5. Lista frekwencyjna dla korpusu porównawczego brytyjskich aktów prawnych

Powyższe wyniki umożliwiają określenie słów najczęstszych i porównanie dystrybucji słów w tekstach tłumaczonych i nietłumaczonych oraz wstępną selekcję elementów do analiz szczegółowych.

3.2. Listy słów kluczowych

Dla omawianych korpusów utworzono również listy słów kluczowych poprzez porównanie list frekwencyjnych dla korpusów równoległych z listami frekwencyjnymi dla korpusów porównawczych. Za słowa kluczowe uznaje się słowa o najbardziej nietypowej frekwencji w zestawieniu z korpusem porównawczym. Lista słów kluczowych zawiera słowa uszeregowane od najbardziej do najmniej nietypowych¹⁶. W tabeli 6. oraz w tabeli 7. przedstawiono słowa kluczowe dla podkorpusów równoległych w zestawieniu z odpowiednimi korpusami porównawczymi.

Lp	Słowo kluczowe	Liczba w KRPL	% KRPL	Liczba w KPPL	Lp	Słowo kluczowe	Liczba w KRPL	% KRPL	Liczba w KPPL
1	M	552688	1,24	405	51	EUROPEJSKĄ	30193	0,07	256
2	WE	301690	0,68	2927	52	RAMACH	54346	0,12	1165
3	ARTYKUŁ	123113	0,28	8	53	IMIENIU	29092	0,07	245
4	ROZPORZĄ- DZENIE	127015	0,29	541	54	WSPÓLNOTE	19536	0,04	17
5	ZGODNIE	149402	0,34	2066	55	STOSOWANE	24780	0,06	142
6	ŻE	154807	0,35	2650	56	ZASTOSO- WANIE	31982	0,07	357
7	KOMISJI	136219	0,31	2727	57	JAKO	52006	0,12	1126
8	CZŁONKOW- SKIE	59592	0,13	183	58	MOCY	34708	0,08	460
9	ROZPORZĄ- DZENIA	145945	0,33	3300	59	ZATEM	17587	0,04	2
10	NALEŻY	87873	0,20	1054	60	WSPÓLNOT	24157	0,05	143
11	NINIEJSZE	44333	0,10	2	61	PAŃSTWO	27240	0,06	236
12	NINIEJSZEGO	50672	0,11	166	62	USTANA- WIAJĄCE	20233	0,05	55

16 <http://www.lexically.net/downloads/version6/HTML/index.html?keywords.htm>

Lp	Słowo kluczowe	Liczba w KRPL	% KRPL	Liczba w KPPL	Lp	Słowo kluczowe	Liczba w KRPL	% KRPL	Liczba w KPPL
13	ODNIESIENIU	70692	0,16	762	63	EUROPEJ- SKICH	27233	0,06	241
14	WSPÓLNOTY	45002	0,10	75	64	DOTYCZĄCE	56390	0,13	1373
15	ZAŁĄCZNIKU	52989	0,12	334	65	PRZYWÓZ	19878	0,04	52
16	DLA	182632	0,41	5812	66	PRZYWOZU	20847	0,05	74
17	KOMISJA	70687	0,16	920	67	ROZPORZĄ- DZENIEM	23442	0,05	156
18	SĄ	134635	0,30	3564	68	POWINNY	27633	0,06	289
19	DS	32346	0,07	14	69	PRZYJMUJE	23808	0,05	180
20	JAK	50987	0,11	494	70	TO	64532	0,15	1851
21	EUR	30402	0,07	2	71	USTANA- WIAJĄCY	15389	0,03	2
22	EWG	50001	0,11	482	72	TRAKTATU	20890	0,05	115
23	ABY	39349	0,09	202	73	JEST	187938	0,42	8647
24	UWZGLĘ- DNIAJĄC	60291	0,14	870	74	WSZYSTKIE	24267	0,05	238
25	UWAGI	35230	0,08	116	75	STRONY	33477	0,08	584
26	RADY	109102	0,25	2937	76	PONIEWAŻ	14163	0,03	5
27	PRODUKTÓW	57623	0,13	807	77	WARTOŚCI	41310	0,09	920
28	ZAŁĄCZNIK	39995	0,09	273	78	CO	67191	0,15	2141
29	SPRAWIE	91447	0,21	2237	79	MUSZĄ	18977	0,04	118
30	ŚRODKI	74514	0,17	1606	80	KRAJÓW	15546	0,03	39
31	DECYZJA	42245	0,10	441	81	PAŃSTWACH	21550	0,05	200
32	CELU	88097	0,20	2293	82	CENY	23390	0,05	260
33	CZŁONKOW- SKICH	46032	0,10	586	83	REPUBLIKI	22618	0,05	239
34	JEŚLI	26864	0,06	54	84	KOMITETU	25199	0,06	330
35	DYREKTYWY	43785	0,10	539	85	PRODUKTY	20531	0,05	187
36	TE	58578	0,13	1098	86	EUROPEJ- SKIEGO	40674	0,09	963
37	WSZYSTKICH	39775	0,09	418	87	CELÓW	28933	0,07	483
38	KTÓRE	91382	0,21	2575	88	CUKRU	14570	0,03	46
39	MUSI	26873	0,06	88	89	BYĆ	85994	0,19	3279
40	MIĘDZY	43304	0,10	598	90	PRODUKCJI	26099	0,06	416
41	SPORZĄ- DZONO	21828	0,05	7	91	DO	684939	1,54	41320
42	EOG	22365	0,05	15	92	UWADZE	24947	0,06	384
43	TRAKTAT	21007	0,05	0	93	CZY	27515	0,06	483
44	NINIEJSZYM	25631	0,06	103	94	PARLA- MENTU	28418	0,06	523
45	BRUKSELI	21225	0,05	15	95	BY	14326	0,03	60
46	ZOSTAĆ	33870	0,08	338	96	AKAPIT	12759	0,03	31
47	PAŃSTW	45005	0,10	743	97	TAKIE	20604	0,05	254
48	II	39842	0,09	555	98	NIEKTÓRYCH	32470	0,07	717

Lp	Słowo kluczowe	Liczba w KRPL	% KRPL	Liczba w KPPL	Lp	Słowo kluczowe	Liczba w KRPL	% KRPL	Liczba w KPPL
49	RYNKU	42061	0,09	639	99	WIĄŻE	13146	0,03	42
50	ZAŁĄCZNIKA	31434	0,07	282	100	STOSOWANIA	28838	0,06	581

Tabela 6. Lista słów kluczowych dla podkorpusu polskiego w porównaniu z korpusem porównawczym polskich aktów prawnych

Lp	Słowo kluczowe	Liczba w KRAN	% KRAN	Liczba w KPAN	Lp	Słowo kluczowe	Liczba w KRAN	% KRAN	Liczba w KPAN
1	REGULATION	323248	0,65	3348	51	IMPORTS	26396	0,05	9
2	EC	264468	0,53	337	52	EXPORT	30018	0,06	220
3	SHALL	389849	0,78	13279	53	SPECIFIC	34934	0,07	633
4	Â€	220035	0,44	0	54	TECHNICAL	30485	0,06	285
5	ARTICLE	265127	0,53	3784	55	APPLICABLE	48829	0,10	2211
6	EUROPEAN	197566	0,40	1606	56	HAVING	72047	0,14	5449
7	ANNEX	110035	0,22	128	57	AGREEMENT	75886	0,15	6029
8	STATES	116105	0,23	591	58	WHEREAS	23908	0,05	3
9	COMMISSION	225198	0,45	11814	59	INTENDED	33861	0,07	767
10	NO	206352	0,41	10614	60	IMPLEMENTATION	29983	0,06	438
11	Â	93647	0,19	0	61	ANNEXÂ	23064	0,05	0
12	COMMUNITY	138744	0,28	3795	62	COVER	26576	0,05	192
13	ARTICLEÂ	88134	0,18	0	63	BRUSSELS	22404	0,05	28
14	PRODUCTS	89761	0,18	310	64	ENSURE	33311	0,07	884
15	NOÂ	70215	0,14	0	65	FROM	197610	0,40	28961
16	SHOULD	94760	0,19	1763	66	TYPE	30416	0,06	689
17	DIRECTIVE	77412	0,16	556	67	SYSTEM	44171	0,09	2316
18	THE	4105230	8,25	885054	68	FREE	28422	0,06	578
19	MEASURES	78550	0,16	1278	69	COMMON	34719	0,07	1280
20	UNION	74057	0,15	1179	70	PRICE	28733	0,06	758
21	COUNCIL	130437	0,26	7931	71	ACCORDING	23508	0,05	307
22	DECISION	114810	0,23	5941	72	REPLACED	22704	0,05	259
23	MEMBER	160637	0,32	12369	73	PRICES	23238	0,05	327
24	EEC	51272	0,10	47	74	ITS	111997	0,23	13888
25	DOWN	56087	0,11	355	75	PURSUANT	30002	0,06	992

Lp	Słowo kluczowe	Liczba w KRAN	% KRAN	Liczba w KPAN	Lp	Słowo kluczowe	Liczba w KRAN	% KRAN	Liczba w KPAN
26	THEREOF	42656	0,09	37 76	76	JULY	24340	0,05	440
27	TREATY	46787	0,09	396 77	77	SET	52640	0,11	3945
28	THESE	46434	0,09	491 78	78	SUGAR	17632	0,04	1
29	REGARD	72444	0,15	3127 79	79	OCTOBER	22485	0,05	316
30	AND	1272934	2,56	252007 80	80	PROTO- COL	20449	0,04	204
31	REPUBLIC	39023	0,08	64 81	81	CONSI- DERED	24951	0,05	588
32	REFERRED	77864	0,16	4087 82	82	PRO- DUCTION	32776	0,07	1435
33	MARKET	62329	0,13	2252 83	83	FOLLO- WING	85843	0,17	9577
34	ESTABLI- SHING	39318	0,08	295 84	84	THIRD	35526	0,07	1784
35	IMPORT	34363	0,07	41 85	85	PROVI- DED	67894	0,14	6583
36	PRODUCT	36670	0,07	181 86	86	USED	49706	0,10	3851
37	ON	423319	0,85	69176 87	87	OFFICIAL	28007	0,06	998
38	ADOPTED	37792	0,08	357 88	88	ORIGIN	16427	0,03	19
39	COUNTRIES	35452	0,07	227 89	89	BASIS	42287	0,08	2840
40	THEREFORE	32454	0,07	100 90	90	BE	506839	1,02	99333
41	DATA	43951	0,09	1103 91	91	JOURNAL	16272	0,03	29
42	CONCER- NING	35325	0,07	372 92	92	LAYING	18405	0,04	167
43	DECEMBER	39327	0,08	722 93	93	COMMU- NITIES	26376	0,05	914
44	EU	37983	0,08	619 94	94	DETAILED	15820	0,03	25
45	BETWEEN	61611	0,12	3255 95	95	NUMBER	43019	0,09	3096
46	POINT	36832	0,07	622 96	96	IMPLE- MENTING	18690	0,04	224
47	ALL	84191	0,17	6541 97	97	HOWEVER	21218	0,04	441
48	THEIR	71773	0,14	4756 98	98	VALUES	18620	0,04	223
49	COMPETENT	32168	0,06	323 99	99	ORIGINA- TING	15882	0,03	54
50	JUNE	30972	0,06	249 100	100	ECONO- MIC	28611	0,06	1270

Tabela 7. Lista słów kluczowych dla podkorpusu angielskiego KRAN w porównaniu z korpusem porównawczym brytyjskich aktów prawnych

Lista słów kluczowych umożliwia prześledzenie nietypowych częstotliwości słów w korpusie tekstów tłumaczonych w porównaniu z korpusem nietłumaczonych tekstów prawnych. W podkorpusie tekstów tłumaczonych na język polski zwraca uwagę nadreprezentacja czasownika niewłaściwego *należy* (dzieśnięte miejsce na liście słów kluczowych). Na liście frekwencyjnej dla korpusu

równoległego występuje on na miejscu trzydziestym siódmym, zaś w korpusie porównawczym polskich ustaw pojawia się dopiero na miejscu dwieście siedemdziesiątym piątym. Stąd wniosek, że liczba wystąpień formy *należy* jest skutkiem interferencji tekstu źródłowego i prawdopodobnie odpowiada angielskiemu *shall* i *should*. Dokładne badania korpusu równoległego wyrównanego mogłyby wykazać, w jakim stopniu nadreprezentacja formy *należy* jest rzeczywiście wynikiem tłumaczenia wymienionych czasowników modalnych. Na liście słów kluczowych dla podkorpusu angielskiego KRAN czasownik modalny *shall* zajmuje trzecie miejsce. Lista frekwencyjna dla tego podkorpusu podaje *shall* na dwunastym miejscu, zaś lista frekwencyjna dla korpusu aktów brytyjskich na miejscu sto trzecim. Czasownik modalny *should* również pojawia się wśród pierwszych dwudziestu słów kluczowych dla podkorpusu angielskiego. W samym podkorpusie KRAN tekstów prawnych UE występuje on na miejscu czterdziestym piątym, a w korpusie porównawczym aktów brytyjskich KPAN dopiero na miejscu siedemset trzydziestym siódmym. Wyrównanie porównanie tekstów źródłowych z docelowymi korpusu równoległego pokazuje, że wysokie miejsce *należy* nie jest skutkiem tłumaczenia czasownika modalnego *shall*. W dalszych badaniach konieczna jest analiza jakościowa tekstów wyrównanych źródłowych i docelowych w celu zbadania kontekstu wymienionych słów i uściślenia wyników badań ilościowych.

Ze względu na założony cel badawczy, jakim jest weryfikacja hipotezy o odmienności gramatycznej języka tłumaczonego, za pomocą korpusu równoległego i porównawczych zbadano również wstępnie dystrybucję form nieosobowych zakończonych na *-no* oraz *-to*, czyli bezosobników (za Saloni, Świdziński, 2007: 150). Tabela 8. przedstawia dwadzieścia najczęstszych form bezosobników zakończonych na *-no* oraz dwadzieścia zakończonych na *-to* oraz liczbę okazji danej formy w korpusie tłumaczonym i korpusie porównawczym.

Typ	KRPL		KPPL				
	Liczba okazów	Typ	Liczba okazów	Typ	Liczba okazów		
sporządzono	21828	przyjęto	1437	wydano	127	zawarto	44
określono	5336	podjęto	781	stwierdzono	71	wszczęto	30
stwierdzono	4345	zawarto	439	udzielono	66	rozpoczęto	18
ustalono	3627	wykryto	397	dokonano	61	cofnięto	16
uznano	2720	osiągnięto	314	przyznano	42	podjęto	10
przyznano	2033	oparto	266	orzeczone	41	usunięto	7
przedstawiono	2000	wzięto	232	Nadano	36	otwarto	6
przewidziano	1799	użyto	227	powierzono	32	wytknięto	4
dokonano	1797	usunięto	198	wniesiono	29	zajęto	4
podano	1771	wszczęto	191	złożono	29	oparto	3

Typ	KRPL		KPPL				
	Liczba okazów	Typ	Liczba okazów	Typ	Liczba okazów		
złożono	1733	rozpoczęto	187	ustalono	25	przyjęto	3
wskazano	1498	Ujęto	178	uzyskano	24	ujęto	3
ustanowiono	1332	wysunięto	131	przeprowadzono	22	dożyto	2
przeprowadzono	1298	objęto	109	zastosowano	22	objęto	2
udzielono	1195	otwarto	109	określono	21	pominięto	2
zastosowano	1178	cofnięto	66	przekazano	21	powzięto	2
opisano	1143	wyciągnięto	49	odmówiono	19	wykryto	2
wykazano	1135	zamknięto	46	postawiono	19	zasięgnięto	2
uwzględniono	937	zasięgnięto	40	ogłoszono	18	dotarto	1
wprowadzono	929	odkryto	36	ustanowiono	14	nabyto	1

Tabela 8. Najczęstsze formy nieosobowe wraz z liczbą ich wystąpień w polskim podkorpusie równoległym i w polskim korpusie porównawczym

Przed rozpoczęciem badania założono zgodnie z hipotezą o niedoreprezentacji form typowych dla języka docelowego, że liczba typów form bezosobników będzie dużo niższa w korpusie równoległym w stosunku do liczby typów w korpusie porównawczym. Okazało się jednak, że według wstępnych badań korpus tekstów tłumaczonych wykazuje znaczną różnicę w liczbie typów, lecz podobną wartość procentową liczby typów form bezosobowych do ogólnej liczby typów w korpusie, jak korpus porównawczy. Te wartości wymagają jednak dokładniejszych analiz, gdyż ze względu na przyjęcie do badania korpusów bez dokładniejszego oczyszczenia zbędnych elementów (na przykład słów z języków obcych), dane procentowe mogą się znacząco zmienić.

Formy nieosobowe czasownika	Liczba typów w KRPL	%	Liczba typów w KPPL	%
zakończone na <i>-no</i>	1036	0,33	185	0,34
zakończone na <i>-to</i>	61	0,02	28	0,05

Tabela 9. Łączna liczba typów form bezosobników dla korpusu tekstów tłumaczonych oraz dla korpusu porównawczego

4. Podsumowanie

Przedstawione korpusy równoległe i porównawcze otwierają możliwość weryfikacji wielu hipotez dotyczących języka tłumaczeń i określenia jego cech. Celem kompilacji korpusów wymienionych w niniejszym artykule jest weryfikacja hipotezy o odmienności gramatycznej i leksykalnej tekstów tłumaczonych. Wstępne badania wykazują inną dystrybucję słów w tekstach tłumaczonych w porównaniu z tekstami nietłumaczonymi, co zostało szczególnie uwidocznione w liście słów kluczowych. Jednak drugi element weryfikacji hipotezy

o odmienności gramatycznej pokazuje wstępnie, że w tekstach tłumaczonych nie występuje niedoreprezentacja form nieosobowych czasowników, co dowodzi, że angielska strona bierna nie jest tłumaczona przez zastosowanie strony biernej w języku polskim, ale tłumacze odwołują się do form typowych dla polskiej składni, co przeczyłoby postawionej hipotezie. Ze względu na pilotażowy charakter przedstawionych badań, powyższe wnioski wymagają jednak pogłębionych szczegółowych analiz ilościowych i jakościowych.

Bibliografia

- BIEL, Łucja (2014): The textual fit of translated EU law: a corpus-based study of deontic modality. *The Translator* 20(2), <http://www.tandfonline.com/doi/full/10.1080/-13556509.2014.-909675>, (28-10-2014).
- GRABOWSKI, Łukasz (2011): Korpusy dwu-i wielojęzyczne w służbie tłumacza, leksykografa i badacza: poszukiwanie ekwiwalentów przekładowych w świetle hipotez dotyczących istnienia uniwersaliów tłumaczeniowych. W: Wojciech CHLEBDA (red.): *Na tropach tłumaczeń. W poszukiwaniu odpowiedników przekładowych*. Opole: Wydawnictwo Uniwersytetu Opolskiego, 89–112.
- MAURANEN, Anna (2008): Universal Tendencies in Translation. W: Gunilla ANDERMAN, Margaret ROGERS (red.): *Incorporating Corpora. The Linguist and the Translator*. Clevedon: Multilingual Matters.
- OLOHAN, Maeve (2004): *Introducing Corpora in Translation Studies*. London: Routledge.
- PIOTROWSKI, Tadeusz (2003): Językoznawstwo korpusowe – wstęp do problematyki. W: Stanisław GAJDA (red.): *Językoznawstwo w Polsce. Stan i perspektywy*. Opole: Wydawnictwo Uniwersytetu Opolskiego, 143–154.
- SALONI, Zygmunt, ŚWIDZIŃSKI Marek (1998): *Składnia współczesnego języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- STEINBERGER Ralf, EISELE Andreas, KLOCEK, Szymon, PILOS, Spyridon, SCHLÜTER, Patrick (2012): DGT-TM: A freely Available Translation Memory in 22 Languages. W: Nicoletta CALZOLARI, Khalid CHOUKRI, Thierry DECLERCK, Mehmet Uğur DOĞAN, Bente MAEGAARD, Joseph MARIANI, Asuncion MORENO, Jan ODIJK, Stelios PIPERIDIS (red.): *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*. Istanbul: European Language Resources Association (ELRA), 454–459.