

Anna Bączkowska
Uniwersytet Mikołaja Kopernika w Toruniu,
Wydział Lekarski Collegium Medicum w Bydgoszczy

Korpusy językowe w języku specjalistycznym medycznym jako narzędzie wspomagające przekład i nauczanie języka medycznego

Wstęp

Celem niniejszego artykułu jest omówienie sposobów wykorzystania korpusów językowych w przekładzie medycznym oraz w dydaktyce języka medycznego zarówno na studiach medycznych, jak i filologicznych (w tym w kształceniu podyplomowym tłumaczy medycznych). W pierwszej części artykułu przedstawione zostaną propozycje zmian w kształceniu tłumaczy medycznych oraz kilka refleksji na temat edukacji przyszłych lekarzy w zakresie kompetencji językowych w medycznym języku angielskim. Po tej dyskusji teoretycznej – w drugiej części artykułu – zaprezentowane zostaną wybrane oprogramowania oferowane przez językoznawstwo korpusowe i komputerowe, które mogą być wykorzystane przez tłumacza medycznego, dydaktyka języka medycznego czy też przez autorów naukowych tekstów medycznych.

Na wstępie trzeba stwierdzić, że spośród wielu odmian języka specjalistycznego (technicznego, prawniczego, akademickiego, hotelarskiego itp.), język medyczny jest z pewnością jednym z trudniejszych do opanowania. Wynika to z faktu, że dyskurs medyczny charakteryzuje się dużym nasyceniem terminów fachowych. Terminy te to zwykle wyrazy rzadko używane poza dyskursem medycznym, powstały one bowiem głównie na bazie łaciny i greki

i w takiej też formie w dużym stopniu profesjolekty funkcjonują w obecnej angielszczyźnie medycznej. Jeśli chodzi o inne cechy dyskursu naukowego medycznego, to trzeba wspomnieć też o typowym dla tekstów medycznych neutralnym, bezosobowym tonie, dbałości o precyzję wypowiedzi, a także o częstym użyciu strony biernej, złożonej składni, skondensowanym i oszczędnym stylu, wysokiej frekwencji skrótów tworzonych *ad hoc*, skrótowców i akronimów.

Język medyczny został już opisany przez wielu autorów. Aspekty językowe i terminologiczne omówili m.in. tacy autorzy, jak Elliot Mishler (1985), Henri van Hoof (1998a), Andrew Hutton (2002), Melodie Hull (2010), Joanna Ciecierska i Barbara Jenike (2015), Ewa Kościółkowska-Okońska (2015, 2017); językiem komunikacji klinicznej zajmowali się m.in. Debra Roter i Judith Hall (2006) czy Peih-ying Lu i John Corbett (2012), a procesem przekładu medycznego – Vincent Montalt i Maria Gonzalez Davies (2007). Elementy stylistyki angielskich tekstów medycznych (głównie tekstów farmaceutycznych oraz dokumentacji medycznej) w oparciu o metodologię językoznawstwa korpusowego analizowali m.in. Daniel Paiva (2000) i Łukasz Grabowski (2014, 2015). Przeprowadzono również analizy porównawcze. Maurizio Gotti (2016) porównywał stylistykę angielskiego i włoskiego akademickiego języka angielskiego używanego w tekstach medycznych publikowanych w czasopismach naukowych. W swoich interesujących badaniach zauważył on, że angielski język medyczny cechuje użycie strategii argumentacyjnych mających na celu łagodzenie kategoryczności wypowiedzi. Stąd w tekstach pisanych po angielsku można zauważyć użycie modalności, zwłaszcza czasowników modalnych (*may, would*), czasowników i przymiotników sugerujących niepewność (*appear, suggest; likely*), co nie występuje we włoskich tekstach medycznych (Gotti 2016). Van Hoof (1998b) z kolei porównał medyczny język angielski z medycznym językiem francuskim na poziomie leksyki, w odniesieniu do łacińskich i greckich wyrazów, od których w dużej mierze pochodzą terminy medyczne. Doszedł on do wniosku, że angielskie i francuskie terminy medyczne są często mylące z uwagi na istnienie wielu „fałszywych przyjaciół” (*false friends*, tj. wyrazów takich, jak angielskie *eventually*, które nie oznacza wcale „ewentualnie”, tylko „w końcu”, czy *sympathetic* nieoznacza „sympatyczny”, tylko „współczujący” itp.). Jack Segura (1998) przedstawia swoje refleksje dotyczące tłumaczenia terminologii medycznej z języka angielskiego na język hiszpański w pracach leksykograficznych. Jego zdaniem bilingwalne słowniki medyczne wydawane w Stanach Zjed-

noczonych nie uwzględniają rodzimych odpowiedników w języku hiszpańskim, co prowadzi do upowszechniania się terminów pochodzenia angielskiego.

Edukacja w zakresie medycznego języka angielskiego

Jak wiadomo, aby prawidłowo przetłumaczyć tekst, należy go wpiern zrozumieć. Tłumacz to lingwista, znający język, ale nie lekarz znający jednostki chorobowe i procedury leczenia, zatem zrozumienie tekstu medycznego stanowi dla tłumacza dużą trudność. Nie będąc lekarzem, nie jest on w stanie w pełni samodzielnie oddać sensu tekstu źródłowego nasyconego wysokospecjalistyczną terminologią, dlatego też tłumaczenia medyczne często wymagają konsultacji z lekarzami danej specjalizacji. Trudno wymagać od lingwisty znajomości medycyny, zatem powstaje pytanie: czy kształcenie tłumaczy tekstów medycznych ma w ogóle uzasadnienie?

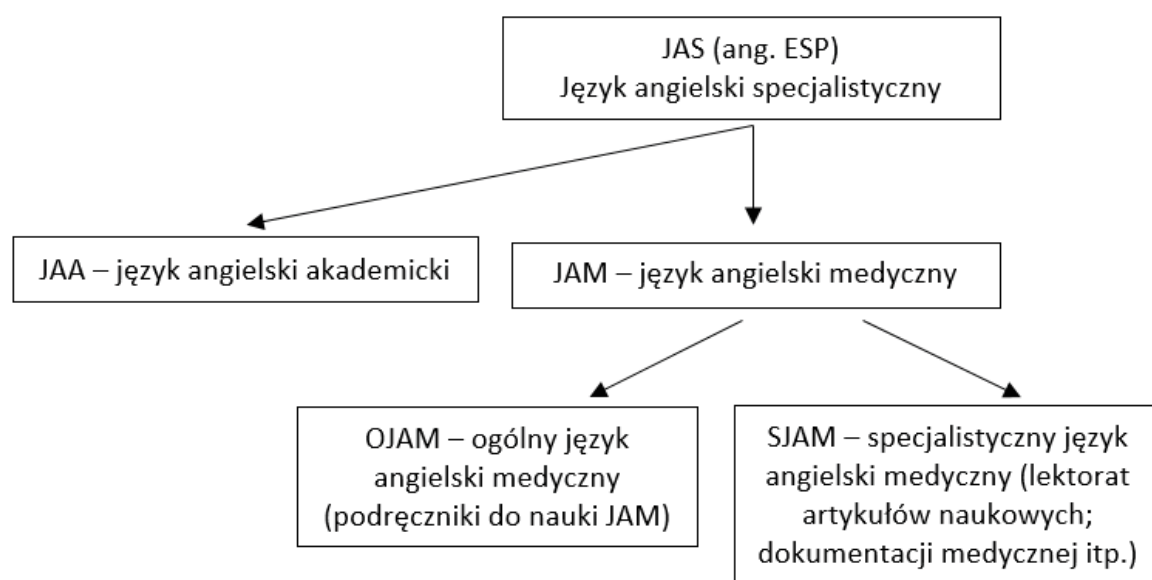
Wydaje się, że – w przeciwieństwie do niektórych odmian specjalistycznego języka angielskiego (np. dotyczącego fryzjerstwa, spawalnictwa, ogrodnictwa, politologii, filozofii, pedagogiki, psychologii), gdzie lingwista po wdrożeniu się w daną dziedzinę jest w stanie dokonać prawidłowego przekładu tekstu – są takie dziedziny specjalistycznego języka angielskiego, w których tłumacz ma niewielkie szanse wykazać się prawidłowym przekładem. Takimi dziedzinami są zapewne zaawansowane teksty techniczne (z zakresu mechaniki, mechatroniki, elektroniki, informatyki itp.). W przypadku tekstów medycznych natomiast zwykły filolog, nawet po specjalizacji/profilu przekładu medycznego, w starciu z tekstami medycznymi musi się zmierzyć z wieloma wyzwaniami.

Patrząc na problem z drugiej strony, można by się zastanowić, czy teksty medyczne powinny być w ogóle tłumaczone, czy raczej nie powinny być pisane w języku angielskim przez autorów tekstów, tj. lekarzy-naukowców. Merytorycznie z dużym prawdopodobieństwem tekst nie zawierałby uchybień, jednak uwzględniając poziom języka angielskiego uzyskiwany przez studentów medycyny podczas lektoratów oferowanych w trakcie studiów medycznych, można mieć poważne wątpliwości, czy lekarz jest w stanie napisać tekst z zakresu medycyny w języku angielskim. Student medycyny (lub lekarz) powinien być co najmniej na poziomie C1 według skali CEFR, tymczasem po zakończeniu lektoratu oczekuje się od studenta znajomości języka obcego na niższym poziomie B2, który i tak nie zawsze jest osiągany przez studentów.

Pytanie, czy to lingwista posiadający szeroką wiedzę medyczną powinien tłumaczyć teksty medyczne, czy raczej lekarz bardzo dobrze znający język angielski, zadawała już sobie Marla O'Neil (1998). Na podstawie swoich badań ankietowych i analizy ich wyników, tj. odpowiedzi udzielonych zarówno przez lekarzy-tłumaczy medycznych, jak i lingwistów-tłumaczy medycznych, doszła ona ostatecznie do wniosku, że i jedni i drudzy mogą być dobrymi tłumaczami tekstów medycznych. W obliczu jednak powyższych rozważań w kontekście polskim wydaje się, że ani tłumacz-lingwista ani autor-lekarz (znający język angielski co najwyżej na poziomie B2) nie są odpowiednio przygotowani do takiego przekładu, zatem wydaje się, że nie powinni oni zajmować się przekładem/pisaniem tekstów medycznych. Można zaproponować kilka rozwiązań w tej sytuacji. Tekst medyczny może być tłumaczony przez zespół co najmniej dwuosobowy, składający się z lingwisty i lekarza o danej specjalizacji. Jest to rozwiązanie najczęściej chyba obecnie stosowane. Drugie możliwe rozwiązanie polegałoby na tym, że tekst pisany jest przez samego lekarza, który zna język angielski na poziomie co najmniej C1 lub wyższym (który ewentualnie przekazuje tekst do sprawdzenia tłumaczowi). Kolejnym wyjściem byłoby tłumaczenie tekstu przez lingwistę, który kończy studia podyplomowe z zaawansowanego przekładu medycznego.

Drugie rozwiązanie zaproponowane powyżej wydaje się potencjalnie osiągać lepsze skutki, jednak wymagałoby wprowadzenia zmian w prowadzeniu zajęć z języka angielskiego na poziomie studiów lekarskich, tj. zwiększenie liczby godzin z zakresu języka angielskiego specjalistycznego (JAS), przy czym przez język specjalistyczny rozumiem nie tylko język angielski medyczny (JAM), oferowany przez podręczniki do nauki języka angielskiego dla studentów medycyny (np. Hutton 2002, Hull 2010, Ciecierska i Jenike 2015), który jest ogólnym językiem angielskim medycznym (dalej OJAM), ale też język angielski akademicki (JAA), aby nauczyć już studentów medycyny, jak pisać teksty medyczne planowane do publikacji w czasopismach naukowych. Skuteczną formą nauczania języka angielskiego specjalistycznego medycznego wydaje się też być wprowadzenie w ramach modułu JAM zajęć poświęconych czytaniu oryginalnych artykułów medycznych i ich tłumaczenie w trakcie zajęć, przy czym preferowane byłyby teksty o tematyce, którą student medycyny zrealizował już w ramach studiów medycznych na I czy na II roku, tj. np. dotyczące anatomii (teksty z zakresu ortopedii), cytologii (teksty z zakresu np. ginekologii), patomorfologii (teksty z zakresu np. onkologii) czy embriologii (teksty z zakresu gi-

nekologii i radiologii). Ten moduł zawęzałby JAM do tekstów z zakresu określonych specjalizacji, byłby to etap ostatni w nauczaniu języka angielskiego w ramach lektoratu – specjalizacyjny język angielski medyczny (SJAM). Ponadto, oprócz tekstów artykułów naukowych z danej tematyki, należałoby również nauczać pisania (i rozumienia) dokumentacji medycznej, np. opis badań obrazowych, epikryz, wypisów ze szpitali itp.



Wydaje się też, że zwiększenie liczby godzin z języka angielskiego specjalistycznego powinno być wprowadzone również na studiach doktoranckich, gdzie oprócz JAA można by już prowadzić w ramach specjalistycznego języka angielskiego zajęcia z języka poszczególnych specjalizacji, do których wcześniej przygotowywali się nauczyciele angiści specjalizujący się w danej specjalizacji medycznej (lub kilku z nich). Dla przykładu, można by prowadzić oddzielne zajęcia specjalistycznego języka angielskiego z zakresu onkologii, okulistyki, laryngologii, ginekologii, diabetologii, radiologii itp.

Trzecie rozwiązanie wspomniane powyżej to studia podyplomowe dla tłumaczy medycznych. W założeniu, w ramach zajęć słuchacze ucą się nie tylko ogólnego języka angielskiego medycznego (tj. treści przekazywane studentom medycyny w trakcie lektoratów) oraz języka angielskiego medycznego z danej specjalizacji medycznej (nauczanego przez lingwistę), ale też uczęszczają na zajęcia z zakresu podstaw medycyny, prowadzone przez wykładowców-lekarzy o różnych specjalizacjach, którzy przekazują wybrane zagadnienia (np. najczęstsze jednostki chorobowe i protokoły postępowania w ich leczeniu) na temat danego działu medycyny. Podobne rozwiązania stosuje się obecnie na

niektórych studiach podyplomowych dla tłumaczy tekstów prawnych przygotowujących do zdania egzaminu państwowego na tłumacza przysięgłego, bowiem teksty obowiązujące na ww. egzaminach są tak daleko specjalistyczne, że aby je zrozumieć i prawidłowo przetłumaczyć, najlepiej byłoby być po studiach prawniczych i znać język angielski na poziomie zaawansowanym. Angliści specjalizujący się w języku prawnym i prawniczym mają z takimi przekładami o wiele więcej problemów niż prawnicy bardzo dobrze znający język angielski. Wzorem takich studiów z przekładu prawniczego i prawnego, można by utworzyć podobny program zajęć prowadzony przez specjalistów zarówno od przekładu jak i medycyny na studiach z przekładu medycznego.

Wzorem też przekładu prawnego i prawniczego (gdzie ten pierwszy oznacza tłumaczenie tekstów prawnych, np. aktów prawnych czy kodeksów, natomiast ten drugi - tłumaczenie tekstów dotyczących prawa), można by się zastanowić, czy nie miałyby uzasadnienia rozróżnienie tekstów medycznych od tekstów medycyny. Przez teksty medyczne rozumiem takie teksty, które zawierają terminologię medyczną o różnym stopniu nasycenia terminami specjalistycznymi, jednak będące na tyle zrozumiałe, że tłumacz, przy użyciu słownika, jest w stanie je przetłumaczyć. Takie teksty napisane są tak, aby osoby interesujące się daną dziedziną, ale niebędące lekarzami, mogły je zrozumieć (np. czasopismo nienaukowe poświęcone zdrowiu, ulotki dla pacjentów). Przez teksty medycyny rozumiem teksty napisane (lub wygłoszone, np. na sympozjach naukowych) przez lekarzy i skierowane do lekarzy, zatem używane do komunikacji z innymi specjalistami (np. epikryzy, opisy badań obrazowych TK, USG itp., dokumentacje badań klinicznych, dokumentacje szpitalne). Wydaje się, że w trakcie lektoratów na kierunku lekarskim, medycznych studiach doktoranckich oraz na podyplomowych studiach z przekładu medycznego naucza się przede wszystkim ogólnego języka medycznego, natomiast do prawidłowego tłumaczenia zaawansowanych tekstów medycznych (tj. nie tekstów o medycynie czy tekstów z podręczników służących edukacji w zakresie języka medycznego) potrzebna jest bardziej zaawansowana wiedza, tj. znajomość języka medycyny (zatem również w jakimś stopniu wiedza z zakresu medycyny).

Ponieważ jednym ze sposobów nauczania się języka wybranej wiedzy z zakresu medycyny jest czytanie, analiza i tłumaczenie oryginalnych, naukowych tekstów medycznych, w dalszej części artykułu zostanie przedstawionych kilka przykładów ćwiczeń, które do nauki języka medycyny oraz do kształcenia tłumaczy specjalistycznych w zakresie medycyny wykorzystują

teksty medyczne opublikowane w medycznych czasopismach naukowych, które stanowią swoistą bazę danych, czyli korpus tekstów medycznych.

Zaproponowana metoda nauczania języka medycznego i kształcenia tłumaczy wpisuje się zatem z jednej strony w tzw. korpusowe badania stosowane (ang. *corpus-based applied studies*), które wykorzystują autentyczne teksty specjalistyczne w opracowywaniu materiałów dydaktycznych i w praktyce tłumaczy (Laviosa 2010: 84), a z drugiej strony w tworzenie korpusów dla potrzeb przekładoznawczych (ang. *translation-driven corpora*), tj. tworzonych i wykorzystywanych przez tłumaczy (Zanettin 2012: 8).

Korpusy językowe

Korpusy językowe, czyli bazy danych tekstowych, to uporządkowane (zwykle ustrukturyzowane) i indeksowane (zawierające kody dotyczące np. oznaczeń morfosyntaktycznych) teksty w formacie elektronicznym. Dzięki kodowaniu możliwe jest ich bardzo szybkie przeszukiwanie za pomocą programów komputerowych z wbudowanymi narzędziami do analizy plików tekstowych (więcej na ten temat np. w Lewandowska-Tomaszczyk 1996, Pęzik 2013, Bączkowska 2016). Korpusy można wykorzystać do efektywnej i wiarygodnej analizy języka różnych typów dyskursu (por. na przykład artykuły dotyczące analizy języka mediów za pomocą korpusów językowych (Baker 2006, Baker i in. 2008; Bednarek 2006; Bączkowska 2016, 2017, w druku a, b), analizy korpusowej tekstów farmaceutycznych i paramedycznych (Lu i Corbett 2012; Grabowski 2014, 2015) czy tekstów prawnych (Goźdź-Roszkowski 2011; Biel 2010, 2014).

Narzędzia korpusowe w analizie języka medycznego

Proste wyszukiwanie w korpusie, tj. pojedynczych wyrazów (np. *ból*) lub ich kolokacji (*ból pleców*, *rwący ból*) czy też konkretnych struktur gramatycznych (np. *regret to say* vs. *regret saying*, „z przykrością zawiadamiać” vs. „żałować, że coś się powiedziało”), nie wymaga użycia szczególnych narzędzi. Inaczej ma się sytuacja, kiedy chcemy wyszukać klasy wyrazów i klasy typów struktur składniowych w korpusach językowych, czyli na przykład różnych przymiotników występujących do trzech miejsc przed danym wyrazem (np. przymiotnik w prepozycji + *ból*), albo różnych czasowników występujących w postpozycji w stosunku do konkretnego czasownika w bezokoliczniku

oraz w gerundium (np. *regret* + bezokolicznik lub gerundium). Wówczas potrzebne już są bardziej zaawansowane sposoby wyszukiwania danych. Do tego celu służą tzw. wyrażenia regularne (regex, skrót od *regular expression*) wykorzystywane w informatyce, językoznawstwie komputerowym i korpusowym.

Wyrażenia regularne. Wyrażenia regularne pozwalają na wyszukiwanie rozmaitych struktur składniowych, np. użycia niektórych wyrazów w stronie czynnej lub biernej. Poniżej przedstawiono kilka przykładów wyrazu *patient* („pacjent”) użytego w stronie biernej dowolnych czasowników, dzięki czemu można szybko znaleźć typowe leksemy łączące się z tym wyrazem. Interesujące jest to, że tłumaczenie niektórych zwrotów jest niedosłowne. Warto nauczać studentów medycyny i przekładu medycznego oraz autorów tekstów naukowych z zakresu medycyny ekwiwalentów, które nie tylko są poprawne, ale też brzmią naturalnie dla rodzimych użytkowników danego języka i/lub które są typowe dla dyskursu medycznego czy innego profesjolektu. Poniższe przykłady (pochodzące z korpusu czasopism DOAJ, podkorpusu tekstów na temat onkologii utworzonego na potrzeby niniejszego artykułu) trafnie pokazują, że na przykład czasownik *zakwalifikować* może przyjąć szereg odpowiedników w języku angielskim, w zależności od kontekstu, zatem *zakwalifikować do operacji* będzie przetłumaczone jako *render resectable*, ale *zakwalifikować do badań radiograficznych* może już brzmieć *evaluate for radiographic response*. Przy okazji warto zwrócić też uwagę na kolokację *radiographic response*, która w dosłownym tłumaczeniu oznaczałaby *odpowieź radiograficzna*, jednak w tekstach medycznych z zakresu radiologii używa się tej kolokacji jako typowej dla tego języka specjalistycznego, obok literalnego odpowiednika *badanie radiograficzne* (*radiographic examination*). Podobnie czasownik *podzielić* może uzyskać różne odpowiedniki w języku angielskim, np. *plotted* czy *randomized*, w miejsce użycia czasowników o wyższej frekwencji, jakimi są *divide* czy *group*. Mimo iż użycie czasowników *divide* i *group* w tłumaczeniu na język angielski byłoby poprawne i zrozumiałe dla anglojęzycznego odbiorcy tekstu medycznego, użycie czasowników rzadszych, o niższej charakterystyce frekwencyjnej, świadczy o znajomości terminologii danego języka specjalistycznego przez autora czy tłumacza. Dyskurs medyczny ma bowiem swoje cechy charakterystyczne wspomniane już wyżej, jedną z najbardziej ewidentnych jest właśnie terminologia fachowa oraz kolokacje. Znajomość ww. kolokacji specjalistycznych pozwala zrozumieć tekst czytany przez specjalistów i tłumaczy, dlatego też należy zaznajamiać tłumaczy tekstów medycznych

oraz specjalistów z danej dziedziny nauk medycznych (najlepiej już na poziomie studiów) z językiem specjalistycznym.

Wybrane przykłady z rzeczownikiem *patients*:

Patients are rendered resectable – pacjenci zostali zakwalifikowani do operacji
Patients were evaluated for radiographic response – pacjenci zostali zakwalifikowani do badań radiograficznych
Patients were plotted into two groups – pacjenci zostali podzieleni na dwie grupy
Patients were randomized to treatment – pacjenci zostali losowo podzieleni na grupy w badaniu

Nawet niewielki korpus, utworzony dla celów demonstracyjnych z dziedziny diabetologii (Diab), liczący ok. 15 tysięcy słów i składający się z zaledwie 5 artykułów naukowych, pozwala zauważyć tendencje słów do współwystępowania z innymi typowymi dla danej dziedziny medycyny. Za pomocą wyrażeń regularnych zbadano występowanie słowa *diabetes* („cukrzyca”) z przymiotnikami w postpozycji. Poniższe przykłady pokazują, że ekwiwalenty kolokacji nie są oczywiste i mogą znacznie różnić się od potencjalnych odpowiedników, które mogłyby się nasuwać w pierwszej kolejności. Jak widać, cukrzyca dziecięca to *juvenile diabetes* (nie *child ~), cukrzyca ciężarnych to *gestational diabetes* (nie *pregnancy related ~, *~ of pregnant women itp.), insulinoniezależna to *noninsulin-dependent* (nie *insuline nondependent), cukrzyca dorosłych to *adult-onset diabetes* (nie *diabetes of adults).

Analiza kolokacji i struktur gramatycznych w tekstach medycznych za pomocą wyrażeń regularnych może okazać się przydatna zarówno dla studentów lektoratów na studiach medycznych, jak i ich wykładowców, którzy pragną nauczać zaawansowanego języka medycznego. Posługując się taką metodą wyszukiwania leksyki w naukowych artykułach medycznych mogą oni przygotować wąskoprofilowe zajęcia poświęcone konkretnej tematyce z zakresu różnych specjalizacji medycznych. Regex jest naturalnie również niezwykle przydatny tłumaczom tekstów medycznych.

Automatyczna ekstrakcja terminów. Na podstawie korpusu Diab można też utworzyć listę najczęściej występujących terminów jedno- lub wielowyrazowych, w tym akronimów. Jak widać na rycinie poniżej, typowe akronimy używane w tekstach o tematyce diabetologicznej to: T1D (cukrzyca typu pierwszego), T2D lub T2DM (cukrzyca typu drugiego), DKA (cukrzycowa kwa-

sica ketonowa), CDVCs (przewlekłe choroby naczyniowe u chorych na cukrzycę), HbA1c (hemoglobina glikowana, frakcja c) itd. Można pominąć terminy stworzone przez autorów okazjonalnie, *ad hoc* – dla potrzeb danego artykułu, np. SGC oznaczający grupę kontrolną o standardowym poziomie glikemicznym (standard glicemic control group) czy IGC oznaczający grupę eksperymentalną w tym samym badaniu (intensive glicemic group). Warto jednak zwrócić uwagę na fakt, że w tekstach medycznych publikowanych w czasopiśmie naukowym istnieje tendencja do tworzenia akronimów występujących tylko w danym tekście (skrót *ad hoc*), z pewnością z uwagi na oszczędność słowa i konkretność – są to cechy typowe dla akademickich tekstów medycznych. Stosowanie takich skrótów jest istotną wskazówką dla tłumacza. Pojawiają się na tej liście również terminy jednowyrazowe związane z tematyką cukrzycy, które dla studentów medycyny nie będą z pewnością niczym nowym, jednak dla tłumacza mogą okazać się pomocne, np. *ketoacidosis* (kwasica ketonowa) czy *nephropathy* (nefropatia). Utworzenie takiej listy terminów przez tłumacza w ramach przygotowań warsztatu swojej pracy może znacząco przyspieszyć późniejszy przekład. Poniższe przykłady zostały wygenerowane za pomocą programu do automatycznej ekstrakcji terminów dostępnego w systemie Sketch Engine.

Oprogramowanie do tworzenia list terminów składających się z dwóch lub więcej wyrazów jest niezwykle przydatne do tworzenia własnych glosariuszy z danej dziedziny. Jak podkreśla van Hoof (1996b: 62), tłumacz powinien być jednocześnie swoim własnym leksykografem, może on wówczas dokonywać częstych aktualizacji potrzebnych mu terminów, wyrażeń, zwrotów i fraz koniecznych do prac translatorycznych w kontekście z jednej strony – ciągle uzupełnianych informacji merytorycznych w medycynie i z drugiej strony – zmieniających się typów zleceń i klientów. Ponieważ terminy przedstawione powyżej ilustrują fragment listy powstałej na skutek automatycznej ekstrakcji danych, nie wszystkie z nich są faktycznie terminami (por. Łukszyn i Zmarzer 2006) medycznymi w sensie dosłownym. Można raczej powiedzieć, że są to ciągi wyrazów, które w danym korpusie najczęściej współwystępowały, zatem z dużą pewnością można ich użyć w tłumaczeniu lub pisząc tekst własny w języku angielskim (np. *all-cause mortality*, *glucose cut-off*, *diabetes prevalence*, *diabetes management*).

Single-word			Multi-word		
	Score	F		Score	F
<input type="checkbox"/> T1D	6,834.75	165	<input type="checkbox"/> diabetes prevalence	1,204.48	27
<input type="checkbox"/> T2DM	2,217.78	54	<input type="checkbox"/> diabetes mellitus	1,006.78	33
<input type="checkbox"/> DKA	1,926.36	46	<input type="checkbox"/> glycemic control	847.89	19
<input type="checkbox"/> Ruijin	1,274.50	29	<input type="checkbox"/> diabetes care	746.06	17
<input type="checkbox"/> Jiaotong	1,219.07	29	<input type="checkbox"/> diabetic ketoacidosis	696.07	16
<input type="checkbox"/> T2D	1,061.30	25	<input type="checkbox"/> present study	495.40	38
<input type="checkbox"/> STEPS	1,041.45	27	<input type="checkbox"/> dutch diabetes	491.31	11
<input type="checkbox"/> CDVCs	1,026.18	23	<input type="checkbox"/> dutch diabetes prevalence	491.31	11
<input type="checkbox"/> HbA1c	824.26	23	<input type="checkbox"/> glucose cut-off	446.73	10
<input type="checkbox"/> ketoacidosis	781.36	20	<input type="checkbox"/> diabetes management	443.19	10
<input type="checkbox"/> Diabetes	774.89	86	<input type="checkbox"/> all-cause mortality	431.62	10
<input type="checkbox"/> mellitus	767.85	34	<input type="checkbox"/> whole blood	397.83	12
<input type="checkbox"/> prevalence	741.92	106	<input type="checkbox"/> blood glucose	393.09	19
<input type="checkbox"/> nephropathy	741.41	21	<input type="checkbox"/> asian type	357.58	8
<input type="checkbox"/> mmol	623.36	19	<input type="checkbox"/> body mass index	342.13	11
<input type="checkbox"/> cerebrovascular	581.17	18	<input type="checkbox"/> mass index	340.00	11
<input type="checkbox"/> retinopathy	564.72	20	<input type="checkbox"/> worsening nephropathy	313.01	7
<input type="checkbox"/> microvascular	541.25	14	<input type="checkbox"/> present analysis	299.82	7
<input type="checkbox"/> Wiley	536.73	29	<input type="checkbox"/> risk ratio	297.26	7
<input type="checkbox"/> Tonga	526.98	20	<input type="checkbox"/> body mass	285.64	11
<input type="checkbox"/> CI	493.74	36	<input type="checkbox"/> plasma glucose	278.49	8
<input type="checkbox"/> IGC	471.23	12	<input type="checkbox"/> classification of diabetes mellitus	268.44	6
<input type="checkbox"/> Diamicon	445.42	10	<input type="checkbox"/> standard glycemic control	268.44	6
<input type="checkbox"/> SGC	428.04	12	<input type="checkbox"/> asian ethnicity	268.44	6
<input type="checkbox"/> IQR	424.03	10	<input type="checkbox"/> glucose meter	268.44	6

Terminologię można w ten sposób ustalić dla dowolnej liczby tekstów i w obrębie wybranej dziedziny medycyny lub kilku dziedzin jednocześnie. Dzięki temu uzyskujemy bardzo szybko dostęp do leksyki specjalistycznej, niezbędnej dla autorów tekstów medycznych, studentów medycyny i lingwistyki stosowanej o specjalności tłumaczeń medycznych, pozwalającej rozumieć artykuły naukowe, pisać je oraz tłumaczyć.

Konkordancje. Kontekst dla każdego z powyższych terminów można rozszerzyć wyświetlając konkordancje, jak zaprezentowano poniżej. To pozwala zauważyć więcej szczegółów, na przykład struktury gramatyczne i jednostki leksykalne, które współwystępują z analizowanym terminem¹:

or parent selfreport. Episodes of *diabetic ketoacidosis* (DKA), defined according to Diabetes Control A, Placzek K, Holl RW. Predictors of *diabetic ketoacidosis* in children and adolescents with type 1 % were underweight. The prevalence of *diabetic ketoacidosis* (DKA) at onset was 50.1%. The proportion of in Chinese T1D patients at diagnosis. *Diabetic ketoacidosis* at diagnosis is not only a life-threatening

Jak wynika z konkordancji, *diabetic ketoacidosis* łączy się z takimi ciągami wyrazów, jak: *the prevalence of* ~ (występowanie kwasicy ketonowej), *predictors of* ~ (czynniki rokujące, wskazujące na), *in children* (u dzieci), *~ at diagnosis* (kwasica ketonowa wykryta podczas diagnozy), czy *episodes of* ~ (epizody kwasicy ketonowej). Niewprawny autor czy tłumacz nieznający powyższych kolokacji mógłby uciec się do innych wyrażen, które teoretycznie, w innym typie dyskursu, miałyby rację bytu, tłumacząc na przykład „występowanie” jako *occurrence*, które również oznacza „występować”, choć nie pociąga za sobą konotacji dotyczącej powszechności występowania jakiegoś stanu czy choroby. Z kolei „czynnik rokujący” to mógłby być teoretycznie *factor* czy *indicator* w połączeniu z czasownikiem *prognosticate*. Znajomość wyrażen występujących w konkordancjach jednak znacząco ułatwia przekład tekstów medycznych oraz ich pisanie, a także nadaje im formę i stylistykę naukowych tekstów medycznych.

Słowa kluczowe. Nie wszystkie słowa istotne dla danej bazy tekstów mogą się pojawić na liście automatycznie wygenerowanych terminów, dlatego w ramach tworzenia warsztatu tłumacza czy autora dobrze jest wygenerować również listę słów kluczowych (Scott i Tribble 2006, rozdział 5). Pozwolą one zapoznać się (studentom, lekarzom-autorom, tłumaczom, wykładowcom) z jednostkami leksykalnymi nie tyle najczęściej występującymi w danych tekstach, ile – co ważniejsze – najbardziej dla nich istotnymi, swoistymi, typowymi. Słowa kluczowe to takie jednostki leksykalne, które są nadreprezentowane w danej bazie tekstowej w stosunku do języka ogólnego. Dla tekstów z bazy Diab zawierającej artykuły z zakresu diabetologii, lista słów kluczowa wygląda następująco:

¹ Zastosowanie konkordancji w nauczaniu języka medycznego opisuje też Laura Gavioli 2005, rozdz. 6.2.

word	frequency	frequency/mill
T1D	165	7354.6
Diabetes	115	5125.9
T2DM	54	2407.0
DKA	46	2050.4
diabetes	201	8959.2
Ruijin	29	1292.6
Jiaotong	29	1292.6
ADVANCE	29	1292.6
T2D	25	1114.3
CDVCs	24	1069.8
HbA1c	23	1025.2
mellitus	34	1515.5
Wiley	29	1292.6
glycemic	19	846.9
ketoacidosis	20	891.5
STEPS	27	1203.5
BMI	20	891.5
ACCORD	17	757.7
nephropathy	21	936.0
RR	23	1025.2
cerebrovascular	16	713.2
retinopathy	20	891.5
SGC	12	534.9
microvascular	14	624.0
Tonga	20	891.5
IGC	12	534.9
RCTs	11	490.3
adult-onset	11	490.3
Guangdong	14	624.0
prevalence	98	4368.2
Vascular	11	490.3
ethnicities	11	490.3
Diamicron	10	445.7
Pretext	10	445.7
LI	13	579.5
IQR	10	445.7
Ethnicity	11	490.3
Shanghai	29	1292.6
glucose	58	2585.2

Słowa kluczowe dla korpusu Diab

Jest to jedynie początek wygenerowanej listy, który zawiera zarówno informację o kluczowości, jak i charakterystykę frekwencji poszczególnych słów kluczowych. Jak łatwo zauważyć, wartości te różnią się od siebie, bowiem słowa, które często występują, nie muszą być słowami charakterystycznymi dla danego typu dyskursu (tj. kluczowymi). Kluczowość oblicza się na podstawie porównania analizowanych tekstów do korpusu referencyjnego, który stanowi punkt odniesienia dla ustaleń, na ile dane słowo jest typowe dla języka danego tekstu (specjalistycznego), a nie jest typowe dla języka ogólnego. Słowa kluczowe wyekscerpowane przed przystąpieniem do tłumaczenia tekstu czy jego pisania z pewnością ułatwią tłumaczowi czy autorowi tekstu wybór jednostek leksykalnych w docelowym języku angielskim (mogą być dołączone do glosariusza utworzonego na potrzeby tłumaczenia danego tekstu). Dydaktyk

nauczający specjalistycznego języka medycznego z kolei może w ten sposób sporządzić listę słówek obowiązujących studentów dla poszczególnych tematów medycznych i w ten sposób lista słów kluczowych może stanowić uzupełnienie listy terminów. Ponadto, słowa kluczowe stanowią zwykle dużą wskazówkę dotyczącą tematyki tekstu (Baker i in. 2008), student zatem może spróbować przewidzieć na podstawie tylko słów kluczowych, o czym jest dany tekst.

Ponieważ teksty w korpusie Diab składają się m.in. z informacji o badaniach przeprowadzonych w szpitalu Ruijin w Szanghaju oraz na wyspach Tonga, informacje te pojawiają się często w bazie Diab i są dla niej istotne, stąd te nazwy własne ujęte są na ww. liście. Nie są one naturalnie istotne dla autorów tekstów czy tłumaczy i należy je pominąć podczas analizy języka medycznego, jednak mogą być przydatne podczas zajęć dydaktycznych prowadzonych na podstawie tekstów oryginalnych artykułów medycznych polegających na streszczeniu lub antycypacji treści artykułów w oparciu o słowa kluczowe.

Wizualizacja kolokacji. Program LancsBox (v.2.0) pozwala wyszukiwać w korpusach (wbudowanych w program oraz własnych) i wizualizować sieć kolokacji z jednym lub kilkoma słowami (Brezina i in. 2015, Gablasova i in. 2017). Poniższy diagram (ryc. 1) ilustruje kolokacje wyrazu *cancer* („rak”) z innymi wyrazami (do trzech w prawym i lewym kontekście, LR3). Dzięki strzałkom łatwo jest odtworzyć kolokacje, w których występowały słowa *cancer* i *endometria* oraz *early-stage* lub utworzyć kolokacje, w których mogą one potencjalnie występować, np. *uterine-confined endometrium*, *endometrial staging*, *early-stage carcinoma*, *surgical endometrial cancer*. Wybór kolokacji oparty jest na częstości współwystępowania badanych słów.

Rycina 1 zawiera niewielką liczbę słów w sieci, bowiem wygenerowana zastała z bazy demonstracyjnej utworzonej na potrzeby niniejszego artykułu i zawierającej jedynie 5 tekstów. Przy większych korpusach można tworzyć bardziej skomplikowane diagramy poprzez określenie większej liczby najczęściej współwystępujących słów. Umożliwia to samodzielne tworzenie z podanych wyrazów fragmentów potencjalnych zdań, które mogły wystąpić w oryginalnych tekstach w skompilowanym korpusie. Dla przykładu, w korpusie artykułów dotyczących cukrzycy zaznaczyłam słowo *cukrzyca* oraz trzy inne często z nim współwystępujące słowa: *incidents*, *patients*, *diagnosed*. Dzięki temu możliwe jest odtworzenie na przykład następujących ciągów wyrazów:

Wizualizację kolokacji można utworzyć z korpusów wbudowanych w program (nie ma jednak korpusu tekstów medycznych, ale jest korpus dotyczący środowiska naturalnego) lub też skompilować własny korpus i wykorzystać jedynie narzędzia z LancsBox do eksploracji danych i generowania wizualizacji słów z dziedziny, która nas interesuje. Można zatem stworzyć korpus na przykład z artykułów medycznych z zakresu laryngologii i sprawdzać powiązania semantyczno-leksykalne charakterystyczne tylko dla tej dziedziny. W trakcie zajęć lektoratu czy zajęć dla studentów/słuchaczy studiów o profilu przekładu medycznego można analizować znaczenia leksykalne i semantyczno-konceptualne wybranych terminów i słów, a także ich najczęstsze kolokacje. Można również poprosić studentów o pogrupowanie poszczególnych słów według części mowy i w ten sposób stworzyć oryginalną listę słów dla konkretnych tekstów. Na podstawie takich list studenci mogą następnie pisać krótkie teksty własne opisujące jakiś przypadek choroby.

Analiza tagów. System Sketch Engine pozwala również sprawdzić (za pomocą narzędzia Node tag) w jakiej formie gramatycznej używane jest dane słowo, czyli jakie ma tagi². Na przykład wyraz *diabetes* („cukrzyca”) występuje w liczbie pojedynczej, rzadko jako nazwa własna (NP), a w liczbie mnogiej występuje tylko w sąsiedztwie innych rzeczowników w tej samej frazie nominalnej (NNS, np. *disorders such as hypothyroidism or type I diabetes have...*).

P N	NN	6,557	
P N	NP	1,170	
P N	NNS	27	

Ryc. 4. Formy gramatyczne dla *diabetes* występujące w korpusie MWC

Natomiast wyraz *cancer* („rak”) występuje najczęściej w liczbie pojedynczej, ale w liczbie mnogiej również jest używany. Ponadto, odnotowano prawie 4 tysiące przypadków użycia tego słowa jako nazwy własnej (nazwy jednostki chorobowej).

P N	NN	18,712	
P N	NP	3,736	
P N	NNS	1,462	

Ryc. 5. Formy gramatyczne dla *cancer* występujące w korpusie MWC

² Poniższe dane pochodzą z korpusu języka medycznego – Medical Web Corpus (dalej MWC), liczącego ok. 42 tysięcy słów.

czaj niekompletne frazy nominalne lub przyimkowe (Biber 1996). Z wygenerowanej listy trigramów wyekscerpowanej z korpusu Diab wynika, że bardzo częstymi zbitkami w analizowanych naukowych tekstach medycznych są: *as well as* („jak również”, mimo że można takie znaczenie wyrazić innymi wyrazami, np. *too* czy *also*), *in the treatment of*, *the presence of*, *due to the*, *a number of*, *in the course of*, *in the absence of*, *the case of*, *it may be*, *as a result*, *a variety of*, itp.

Jak widać na rycinie 7, najwyżej odnotowana jest zbitka *as well as*, która ma funkcję organizacji dyskursu (*discourse organisers*). Podobnie zbitką organizacji dyskursu jest *one of the* uplasowana na drugim miejscu częstości użycia. Zbitką referencyjną (*referential*), odwołująca się do miejsca, jest *in the United States*, do czasu – np. *at the same time*, a do ilości – *there is no*. Zbitki *should not be* czy *can not be* z kolei ilustrują zbitki leksykalne postawy (*stance*), czyli takie, które wyrażają stanowisko autora wobec omawianego zagadnienia.

<u>word (n-grams)</u>	<u>frequency</u>
as well as	9,058
one of the	7,382
part of the	7,193
the use of	6,996
of the body	5,241
the treatment of	5,073
the presence of	4,926
of the disease	4,898
in which the	4,283
can not be	4,229
there is a	4,189
of the heart	4,081
the end of	4,079
the United States	4,070
the fact that	4,028
in order to	3,940
in patients with	3,628
a number of	3,431
there is no	3,427
that it is	3,400
it may be	3,353
due to the	3,312
of the patient	3,279
the development of	3,233
the same time	3,166
some of the	3,150
more or less	3,033
the result of	2,987
the number of	2,958
DISEASES OF THE	2,900
end of the	2,860
it is not	2,857
side of the	2,844
the course of	2,831
should not be	2,818
the other hand	2,803
in the treatment	2,754
a result of	2,734

Ryc. 7. Trigramy zbitok leksykalnych wyekscerpowane z korpusu Diab

N-gramy można ustawić również na 4-wyrazowe, co jest zalecane przez niektórych autorów (np. Goźdź-Roszkowski 2011). Tabela poniżej przedstawia zbitki leksykalne wyekscerpowane z korpusu tekstów onkologicznych opublikowanych przez polskie czasopisma (korpus *Oncol_pl*) oraz wydawanych w Wielkiej Brytanii (korpus *Oncol_GB*)³. Badania porównawcze tego typu pozwalają zaobserwować różnice w użyciu zbitki leksykalnych tekstów medycznych, w tym przypadku z dziedziny onkologii, napisanych przez polskich autorów w porównaniu do autorów brytyjskich lub pochodzących z innych krajów (więcej na ten temat w: Bączkowska, w tym tomie). Korpus *Oncol_GB* stanowi niejako korpus referencyjny dla korpusu *Oncol_pl*. Poniższy diagram przedstawia fragmenty list ze zbitkami leksykalnymi. Z pobieżnej analizy wynika, że najczęściej występujące zbitki w obydwu korpusach są stosunkowo zbieżne pod względem strukturalnym. Zbitki te są również w dużej mierze zbieżne z tymi analizowanymi powyżej i pochodzącymi z korpusu *Diab*. Podobne analizy porównawcze mogą być dokonywane na podstawie tekstów pisanych przez studentów w ramach zajęć poświęconych nauczaniu pisania naukowych tekstów medycznych (nauczania dyskursu akademickiego, specjalistycznego).

<i>Korpus oncol_pl</i>	<i>Korpus oncol_GB</i>
on the basis of	as well as the
in the case of	On the other hand
as well as the	for the treatment
on the other hand	on the basis of
as a result of	has been show to
in the form of	are taken into account
is one of the	in the treatment of
the basis of the	in the estimation of
in the course of	as a result of
at the same time	is one of the
with the use of	equivalent isotropic displacement parameters
in the area of	atomic coordinates and isotropic
one of the most	or equivalent isotropic displacement
the fact that the	in terms of the
the aim of this	are estimated using the
the aim of the	used when they are
in the treatment of	taken into account individually

³ Korpus tekstów wydanych w Polsce liczy około 22 mln słów, natomiast korpus tekstów wydanych w Wielkiej Brytanii zawiera ok. 215 mln słów. Teksty pochodzą z naukowych czasopism medycznych.

Znajomość typowych zbitek leksykalnych używanych w tekstach akademickich jest niezbędna do prawidłowego tłumaczenia i pisania tekstów naukowych, zatem nauczanie języka angielskiego akademickiego (medycznych tekstów naukowych), czyli wspomnianego w pierwszej części artykułu JAA, można oprzeć na analizie zbitek leksykalnych.

Niedawno opracowano też programy dostępne online (<http://writeaway.nlpweb.org>; <http://langtest.jp/awsum/>), które w oparciu o dane korpusowe wspomagają pisanie tekstów naukowych, m.in. uzupełniając zbitkami leksykalnymi tekst pisany online (opracowane przez Atsushi Mizumoto).

Listy słów. Istotnym punktem w rozwoju dyskursu medycznego było opracowanie przez Wang i in. (2008), na podstawie korpusów tekstów medycznych, listy słów medycznych (Medical Word List, MWL) najczęściej występujących w naukowych artykułach medycznych. Lista ta powstała na wzór listy słówek typowych dla dyskursu akademickiego (Academic Word List, AWL) zaproponowanej w 2000 roku przez Coxhead (zob. też Coxhead 2011), która z kolei nawiązywała do GSL (General Service List) opracowanej w latach pięćdziesiątych (West 1953) – listy wskazującej liczbę słów, którą należy znać na poszczególnych poziomach językowych. MWL stanowi cenną wskazówkę dla dydaktyków przy opracowywaniu programów nauczania medycznego języka specjalistycznego na studiach medycznych i/lub na studiach z zakresu przekładu medycznego. Wzorując się na MWL i korzystając z korpusu języka akademickiego reprezentującego różne dyscypliny wiedzy, Simpson-Vlach i Ellis (2010) zaproponowali Academic Formulas List (AFL), tj. listę utartych wyrażen najczęściej używanych w tekstach akademickich. AFL, a także listę słów najczęściej używanych w mówionym języku akademickim opracowaną przez Danga i in. (2017). Mogą być one przydatnym materiałem dla dydaktyków uczących języka akademickiego, w tym medycznego.

Podsumowanie

W artykule przedstawiono kilka narzędzi wykorzystywanych w językoznawstwie korpusowym, które mogą okazać się pomocne w zrozumieniu i pisaniu naukowych tekstów medycznych w języku angielskim. Mogą one być przydatne dla lekarzy i naukowców reprezentujących nauki okołomedyczne, w tłumaczeniu takich tekstów przez lingwistów-tłumaczy, a także w dydaktyce angielskiego języka medycznego na poziomie studiów medycznych (w ramach

lektoratów) i studiów podyplomowych z zakresu przekładu medycznego. Omówiono wykorzystanie wyrażen regularnych w ekstrahowaniu określonych struktur gramatycznych i jednostek leksykalnych, automatyczną ekstrakcję terminów, konkordancje, wizualizację kolokacji, funkcję wyrazów w zdaniu (tagi), zbitki leksykalne oraz listy słów. Ponadto, w artykule zasugerowano potencjalną modyfikację form zajęć prowadzonych dla przyszłych lekarzy w zakresie kompetencji językowych, polegającą na zwiększeniu liczby godzin dydaktycznych przeznaczonych na lektorat języka angielskiego tak, aby umożliwić nauczanie nie tylko ogólnego języka medycznego, ale też zaawansowanego języka medycznego charakterystycznego dla różnych specjalizacji medycznych; rozumienia, pisania lub tłumaczenia dokumentacji medycznej, a także nauczania pisania stylistycznie poprawnych tekstów naukowych.

Bibliografia

- Baker Paul (2006), *Using Corpora in Discourse Analysis*, Bloomsbury.
- Baker Paul, Gabrielatos Costas, KhosraviNik M., Krzyżanowski Michał, McEnery Tony, Wodak Ruth (2008), *A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press*. „Discourse and Society”, nr 19 (3), s. 273-306.
- Bączkowska Anna (2016), *Korpusowa analiza dyskursu związanego z tematyką imigracji w brytyjskiej prasie opiniotwórczej*, „Conversatoria Linguistica”, nr 10, s. 7-25.
- Bączkowska Anna (2017), *Krytyczna analiza dyskursu prawicowo-populistycznego: analiza korpusowa przemówień wyborczych Donalda Trumpa*, w: F. Pierzchalski, B. Rydliński (red.), *Autorytarny populizm w XXI wieku. Krytyczna rekonstrukcja*, Elipsa, s. 181-200.
- Bączkowska Anna [w tym tomie], *Funkcjonalna analiza zbitek leksykalnych w dyskursie medycznym*.
- Bączkowska Anna [w druku a], *Obraz Polaka imigranta w brytyjskiej prasie opiniotwórczej: analiza korpusowa i krytyczna analiza dyskursu*, w: I. Benenowska, W. Czechowski, A. Bączkowska (red.), *Komunikowanie wartości – wartość komunikowania*, Wydawnictwo Uniwersytetu Kazimierza Wielkiego w Bydgoszczy.
- Bączkowska Anna [w druku b], *A corpus-assisted critical discourse analysis of “migrants” and “migration” in the British tabloids and quality papers*, w: B. Lewandowska-Tomaszczyk (red.), *Contacts and Contrasts in Culture and Language*, Springer.
- Bednarek Monika (2006), *Evaluation in Media Discourse: Analysis of a Newspaper Corpus*, Bloomsbury.
- Biber Douglas (1996), *University Discourse*, John Benjamins.

- Biber, Douglas, Susan Conrad i Viviana Cortes (2004), „*If you look at...: Lexical bundles in university teaching and textbooks*”, *Applied Linguistics* 25/3, s. 371-405.
- Biel Łucja (2010), *Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential*, [w:] Heine, C., Engberg, J. (red.) *Reconceptualizing LSP*. Online proceedings of the XVII European LSP Symposium Aarhus.
- Biel Łucja (2014), *Lost in the Eurofog. The Textual Fit of Translated Law*, Peter Lang.
- Brezina Vaclav, McEnery Tony, Wattam Stephen [2015], *Collocations in Context*, „*International Journal of Corpus Linguistics*”, 20/2 s. 139-173.
- Ciecierska Joanna, Jenike Barbara (2015), *English for Medicine*, Wydawnictwo Lekarskie PZWL.
- Coxhead Averil (2000), *A new Academic Word List*, „*TESOL Quarterly*” nr 34, s. 213-238.
- Coxhead Averil (2011), *The Academic Word List 10 years on: research and teaching implications*, 11, „*TESOL Quarterly*”, nr 45, s. 355-362.
- Dang, Thi Ngoc Yen, Coxhead Averil, Webb Stuard (2017), *The Academic Spoken Word List*, „*Language Learning*”, nr 67/4, s. 959-997.
- Gotti Maurizio (2016), *Variations in medical discourse for academic purposes*, w: P. Ordoñez-Lopez, N. Edo-Marza, *Medical Discourse in Professional, Academic and Popular Settings*, *Multilingual Matters*, s. 30-64.
- Gablasova Dana, Brezina Vaclav, McEnery Tony (2017), *Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence*, „*Language Learning*”, nr 67, s. 155- 179.
- Gavioli, Laura (2005), *Exploring Corpora for ESP Learning*, John Benjamins.
- Gotti Maurizio, Sarcevic Susan (2006), *Insights into Specialized Translation*, Peter Lang.
- Goźdź-Roszkowski Stanisław (2011), *Patterns of Linguistic Variation in American Legal English. A Corpus-Based Study*, Peter Lang.
- Grabowski Łukasz (2014), *On lexical bundles in Polish Patient Information Leaflets: a corpus-driven study*, „*Studies in Polish Linguistics*”, tom 9/1, s. 21-43.
- Grabowski Łukasz (2015), *Keywords and lexical bundles within English pharmaceutical discourse: a corpus-driven description*, „*English for Specific Purposes*”, nr 38, s. 23-33.
- Hull Melodie (2010), *Medical English Clear and Simple*, F.A. Davis Company.
- Hutton Andrew N. (2002), *An Introduction to Medical Terminology for Health Care*, Elsevier.
- Hyland Ken (2008), „*As can be seen*”: *lexical bundles and disciplinary variation*, „*English for Specific Purposes*”, nr 27, s. 4-11.
- Kościałkowska-Okońska Ewa (2015), *Tłumacz przysięgły a pułapki tekstów medycznych*, „*Rocznik Przekładoznawczy*”, nr 10, s. 127-138.
- Kościałkowska-Okońska Ewa (2017), *Na ścieżce kariery: tłumacz tekstów medycznych w kontekście wymagań zawodowych*, „*Rocznik Przekładoznawczy*” nr 12, s. 189-199.

- Laviosa Sara (2010), *Corpora*, w: Y. Gambier i L. van Doorslaer (red.) *Handbook of Translation Studies*, t. 1, Amsterdam, str. 80-86.
- Lewandowska-Tomaszczyk Barbara (2005), *Podstawy językoznawstwa korpusowego*, Wydawnictwo Uniwersytetu Łódzkiego.
- Lu Peih-ying, Corbett John (2012), *English in Medical Education*, Multilingual Matters.
- Łukaszyn Jerzy, Zmarzer Wanda (2006), *Teoretyczne podstawy terminologii*, Uniwersytet Warszawski.
- Mishler Elliot (1985), *The Discourse of Medicine: Dialectics of Medical Interviews*, Ablex.
- Montalt Vincent, Gonzalez Davies Maria (2007), *Medical Translation Step by Step*, Routledge.
- O'Neal Marla (1998), *Who makes a better medical translator: the medically knowledgeable linguist or the linguistically knowledgeable medical professional? A physician's perspective*, w: Fischbach, M. (red.), *Translation and Medicine*, John Benjamins.
- Paiva Daniel (2000), *Investigating style in a corpus of pharmaceutical leaflets: results of factor analysis*, w: Proceedings of the Student Workshop of the 38th Annual Meeting of the ACL, Hong Kong, China 1-8.10, s. 52-59.
- Pęzik Piotr (2013), *Paradygmat dystrybucyjny w badaniach frazeologicznych. Powtarzalność, reprodukcja i idiomatyzacja*, w: P. Stalmaszczyk (red.), *Metodologie Językoznawstwa. Ewolucja języka, Ewolucja teorii językoznawczych*, s. 141-160. Wydawnictwo Uniwersytetu Łódzkiego.
- Roter Debra L., Hall Judith A. (2006), *Doctors Talking with Patients/Patients Talking with Doctors*, Praeger.
- Segura Jack (1998), *Some thoughts on the Spanish Language in medicine*, w: M. Fischbach (red.), *Translation and Medicine*, John Benjamins.
- Scott Michael, Tribble Chris (2006), *Textual Patterns*, John Benjamins.
- Simpson-Vlach Rita, Ellis Nick (2010), *An Academic Formulas List: New Methods in Phraseology Research*, „Applied Linguistics”, nr 31(4), s. 487-512.
- Van Hoof Henri (1998a), *The language of medicine: a comparative ministudy of English and French*, w: M. Fischbach (red.), *Translation and Medicine*, John Benjamins, s. 49-65.
- Van Hoof Henri (1998b), *A contribution to the history of medical translation in Japan*, w: M. Fischbach (red.), *Translation and Medicine*, John Benjamins, s. 29-38.
- Wang Jing, Liang Shao-Ian, Ge Guang-chun (2008), *Establishment of a Medical Academic Word List*, „English for Specific Purposes”, nr 27, s. 442-458.
- West Michael (1953), *A General Service List of English Words*, Longman.
- Zanettin Frederico (2012), *Translation-Driven Corpora*, Routledge.

Using language corpora in teaching English for Medical Purposes and in translation of research medical texts

Abstract: The aim of this paper is to discuss the problem of how corpus linguistics can be used in the field of English for medical purposes (EMP), in particular in: teaching medical English, teaching medical translation, assisting translators in rendering medical texts and authors in writing medical academic texts. A number of corpus tools and ways of data analysis are presented, such as: automatic term extraction, data visualisation, collocation and tag analysis, using regular expressions in data retrieval, function of lexical bundles and word lists. On a more general note, some suggestions as regards the didactics of EMP are presented, such as increasing the number of teaching hours devoted specifically to advanced EMP, along with the so-called general EMP, and designing classes devoted to the analysis of academic discourse and the teaching of academic writing.

Keywords: *English for Special Purposes, English for Medical Purposes, English for Academic Purposes, language corpora, teaching medical English, translation of medical English, medical academic discourse*

Słowa kluczowe: *profesjolekt, język medyczny, język angielski dla celów specjalistycznych, korpusy językowe, nauczanie języka medycznego, przekład tekstów medycznych, medyczny dyskurs naukowy*