

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

## Funkcjonalna analiza zbitek leksykalnych w dyskursie medycznym

### Wstęp

Artykuł omawia badania dotyczące występowania jednego z typów leksykalnych jednostek wielowyrazowych, tzw. zbitek leksykalnych (ang. *lexical bundles*, Biber 1996), w naukowych tekstach medycznych. Do tego celu wykorzystano korpus tekstów medycznych oraz dokonano analizy ilościowej i jakościowej w zakresie okurencji zbitek leksykalnych. W pierwszej części artykułu zdefiniowane zostaną zbitki leksykalne oraz omówione badania dotyczące naukowego języka medycznego. W drugiej części przedstawione zostaną wyniki badań korpusowych dotyczących użycia zbitek leksykalnych w artykułach naukowych z zakresu medycyny opublikowanych w Polsce (dalej korpus JMP), w porównaniu z korpusem referencyjnym tekstów medycznych stanowiących podkorpus *British National Corpus*.

### Zbitki językowe i podejście frazeologiczne

Zbitki leksykalne (ang. *lexical bundles*) to według Douglasa Bibera (1996, rozdz. 6), który zaproponował ten termin, ciągi wyrazów mające tendencję do współwystępowania (w sytuacjach komunikacyjnych), które nie są idiomatyczne i zwykle nie stanowią odrębnej jednostki strukturalnej, natomiast często są na styku dwóch różnych struktur leksykalno-gramatycznych. Jeśli stanowią one kompletną jednostkę strukturalną, to są pomostem między

dwiema innymi jednostkami strukturalnymi. Takich typów zbitek leksykalnych (dalej ZL) jest jednak niewiele, jedynie ok. 15% w dyskursie mówionym i ok. 5% w dyskursie pisanym. Przykładowe zbitki leksykalne w języku polskim liczące trzy wyrazy to: *na przykład w, który nazywa się, jak wcześniej wspomniano, nie wiem czy itp.* W języku polskim nie ma jednoznacznego odpowiednika tego terminu, z pewnością z uwagi na stosunkowo nowe pojęcie, jakim jest zbitka leksykalna oraz niewiele badań prowadzonych w Polsce na ten temat (badania takie prowadzą np. Goźdz-Roszkowski 2011, Grabowski 2014, Kopa-czyk 2013, jednak opisują je w języku angielskim). Według Słownika Oxford/PWN, w języku angielskim wyraz *bundle* oznacza paczkę, wiązkę, plik, pakiet. Ekwiwalent „zbitka leksykalna” dla angielskiego terminu *lexical bundle* został zaproponowany przez Piotra Pęczika (2012), który powołując się na prace Bibera definiuje ten termin w odniesieniu do dyskursu mówionego w tekstach korpusowych jako „najczęściej powtarzające się sekwencje wyrazów”. Grabowski (2014:21) unika określenia zbitka leksykalna i zastępuje je terminem „zbitka wielowyrazowa”. Zgodności co do terminologii takich związków nie ma nawet w literaturze anglosaskiej. Na przykład Michael Scott i Christopher Tribble (2006) używają określenia *word clusters*, a Banerjee i Pedersen (2003) oraz William Fletcher (2006: 35) piszą o *n-gramach*. W innych badaniach nad frazeologią angielską podobne frazeologiczne jednostki wielowyrazowe Göran Kjellmer (1987) i Bengt Altenberg (1991, 1998) określali jako powtarzające się serie/kombinacje słów, stosując odpowiednio określenia *recurrent strings* oraz *recurent word combinations*, i właśnie cecha rekurencji jest kluczowa dla badań zbitek leksykalnych i podejścia frazeologicznego.

W polskich badaniach nad frazeologią najbliższe zagadnieniom zbitek leksykalnych wydają się być badania Wojciecha Chlebdy (1991) i jego teoria frazematyki, która podkreśla znaczenie rekurencji fraz (które nazywa frazemami) i ich charakteru odtwórczego. Podejście takie ilustruje tzw. paradygmat syntetyczny we frazeologii, w przeciwieństwie do paradygmatu analitycznego, reprezentowanego na przykład przez Stanisława Skorupkę (1952) czy Wiktora W. Winogradowa (1977)<sup>1</sup>. Założenia frazematyki zbieżne są z neofirthiańskim (Firth 1957) i dystrybucyjnym podejściem do frazeologii, na bazie którego rozwija się językoznawstwo korpusowe, w szczególności tak zwany model *corpus-*

---

<sup>1</sup> Więcej na temat frazeologii w badaniach polskich i rosyjskich językoznawców piszę w: Bączkowska (w druku).

*driven* (w przeciwieństwie do *corpus-based*, zob. Tognini-Bonelli 2001) promujący analizy od szczegółu do ogółu (ang. *bottom-up*).

W ogólnym zarysie można powiedzieć, że John Firth zmienił myślenie o semantyce leksykalnej podważając sens badania znaczenia jako oddzielnego bytu, bez odwoływania się do kontekstu, w którym występują leksemy. Firth (1957: 11) uważa, że znaczenie jest ściśle zależne od kontekstu, w którym dany wyraz się znajduje, w szczególności od innych wyrazów, z którymi zwyczajowo współwystępuje. Podejście neofirthiańskie (Flowerdew 2012) zakłada badanie bardzo dużych baz danych tekstowych i opiera się na kontekstowej teorii znaczenia zaproponowanej przez Firtha (1957). Innymi słowy pozwala ono zwerifikować tezy Firtha w danych empirycznych pozyskiwanych na dużą skalę. W badaniach neofirthiańskich punktem wyjścia jest zwykle leksyka, w szczególności charakterystyka syntagmatyczna wyrazów, która w dalszych etapach analizy dyskutowana jest w kontekście struktur gramatycznych, w których dane wyrazy występują. W ten sposób neofirthiańskie podejście do frazeologii łączy badania nad leksyką z badaniami nad konstrukcjami syntaktycznymi, uważając je za nierozzerwalne. Nie można bowiem badać frazeologii koncentrując się jedynie na lekсыce i nie uwzględniając struktur składniowych, w których one występują (Halliday 1992: 63), co jest szczególnie widoczne w przypadku zbitek leksykalnych. Jak zauważa Lynne Flowerdew (2012: 66), teoria gramatyki systemiczno-funkcjonalnej Hallidaya (1992) również zakłada ścisły związek składni i lekсыki, z tą jednak różnicą, że w gramatyce Hallidaya punktem początkowych analizy są struktury gramatyczne, a w drugiej kolejności wypływające z nich wnioski dotyczące lekсыki. Z kolei w korpusowym podejściu neofirthiańskim, przedstawicielami którego są m.in. John Sinclair (1991), Michael Stubbs (1996), Michael Hoey (2004) czy Patrick Hanks (2004), analiza zaczyna się na poziomie lekсыki (semantyki), a właściwie frazy. Badania korpusowe prowadzone w duchu neofirthiańskim koncentrują się na związkach frazeologicznych, w tym, a może przede wszystkim, na związkach luźnych (używając nomenklatury Stanisława Skorupki 1954) i rekurencyjnych (frazemach w rozumieniu Chlebdy 1991).

Warto wspomnieć w tym miejscu, że założenie o ścisłym związku lekсыki ze składnią, i w konsekwencji przekonaniem, że nie można tych dwóch elementów analizować w oderwaniu od siebie, obecne jest również w innych teoriach językoznawczych, na przykład w gramatyce konstrukcji, rozwijanej przez wielu badaczy, w szczególności przez Adelę Goldberg (2006). Gramatyka konstrukcji podważa arbitralność relacji formy i znaczenia (*form-meaning pairing*)

i zastępuje ją tezą o ciągłości leksykalno-syntaktycznej, gdzie konstrukcje są nieredukcjonistyczne, tj. rozumiane są jako holistyczne zjawiska *Gestalt*. Podejście gramatyki konstrukcji może być weryfikowane poprzez badania korpusowe, co proponują na przykład Anatol Stefanowitch i Stefan Gries (2003) w swoich statystycznych analizach tzw. kolostrukcji (*collostructional analysis*). Podobne tezy o konieczności badań leksyki w powiązaniu ze strukturami syntaktycznymi, w których występują, zaproponował już wcześniej Ronald Langacker w swojej gramatyce kognitywnej (1987).

Z podejściem neofirthiańskim do frazeologii ściśle związana jest semantyka dystrybucyjna. Podwaliny dla semantyki dystrybucyjnej stworzył Zellig Harris (1954: 43), który uważa, że różnice w znaczeniu słów zależą od różnicy w dystrybucji ich kolokatów, tj. od ich rozkładu i frekwencji w kontekście. W podejściu dystrybucyjnym, w które wpisują się badania korpusowe, znaczenie jest emergentne, tj. wyłania się z kontekstów. Znaczenia nabierają nie tylko słowa współwystępujące, ale też ich dane statystyczne, na przykład frekwencja ich okurencji w sąsiedztwie badanego słowa oraz siła ich łączliwości (co mierzy np. parametr statystyczny MI3). W przeciwieństwie jednak do tradycyjnego podejścia do znaczenia w semantyce analitycznej, w semantyce (i frazeologii, zob. Pęzik 2013) dystrybucyjnej, a także w semantyce kognitywnej (zob. Geeraerts i Cuykens 2007) znaczenie jest sumą znaczeń wyrazów i związków frazeologicznych, w które wchodzi z analizowanym słowem tworząc tzw. sieć znaczeń (ang. *network of meanings*).

Wracając do pojęcia zbitek leksykalnych trzeba podkreślić, że jednym z kluczowych parametrów w ich analizie jest określenie progu minimalnej liczby okurencji ZL w danym korpusie. W swoich badaniach nad rejestrem akademickim Biber i in. (1999) przyjął jedynie 10 okurencji na milion słów, jako wymaganą minimalną liczbę wystąpień ZL. W innych badaniach Biber i in. (2004) oraz Douglas Biber i Federica Barbieri (2007) przyjęli aż 40 okurencji na milion słów i, jak twierdzi sam Biber, jest to podejście konserwatywne. Inni autorzy przyjmowali różne wartości minimalnej okurencji ZL, na przykład Ken Hyland (2008a) i Vivian Cortes (2004) zaproponowali wartość 20 ZL na milion słów, Yu-Hua Chen i Paul Baker (2010) przyjęli 25 wystąpień, a w innym badaniu (Chen i Baker 2016) jedynie 4 wystąpienia, z uwagi na małe korpusy danych. Ponadto przyjmuje się (Cortes 2004), że ZL muszą wystąpić w co najmniej pięciu tekstach (Chen i Baker 2016 przyjęli min. 3 teksty dla małych korpusów). Minimalna liczba okurencji jest zatem dość arbitralna i zależy od decyzji badacza wynikającej z wielkości korpusu i typu danych.

Podjęcie Bibera do analizy zbitek leksykalnych jest frekwencyjne i oparte na różnicach w rejestrach, Biber bada bowiem głównie różnice w wystąpieniach ZL w tekstach akademickich w dyskursie mówionym vs. w akademickim dyskursie pisanym. Rita Simpson-Vlach i Nick Ellis (2010:493) z kolei przyjmują parametr wagi asocjacji elementów składowych ZL<sup>2</sup> zamiast frekwencji, co według nich pozwala wyeliminować zbitki zawierające wyrazy gramatyczne o wysokiej frekwencji, które – jak twierdzą ci badacze – fałszują wyniki na listach frekwencyjnych.

ZL podzielono na kilka podtypów analizując je na poziomie strukturalnym (Biber i in. 1999, rozdz. 13.2) oraz funkcjonalnym (Biber i in. 1999, rozdz. 13.2; Biber i in. 2004, Hyland 2008a). Podział strukturalny opisuje ZL pod względem leksykalno-gramatycznym (wyróżniając np. zbitki ilustrujące frazy nominalne, werbalne, przyimkowe itp.). Podział funkcjonalny ZL zaproponowany przez Bibera i in. (2004) jest szeroko stosowany, a na jego podstawie powstała klasyfikacja opracowana przez Hylanda (2008a). Biber i in. (1999) dzielą ZL na trzy podstawowe grupy: zbitki organizujące dyskurs (łącznie fragmenty tekstu w logiczny ciąg; ang. *discourse lexical bundles*), zbitki referencyjne (odnoszące się do czegoś, np. czasu, miejsca, pojęcia, przedmiotu; ang. *referential lexical bundles*) i zbitki wyrażające postawę (tj. ocenę, stosunek autora itp.; ang. *stance lexical bundles*). Zbitki organizujące dyskurs wprowadzają nowy temat, zmieniają go lub informują o zakończeniu, ale także wyjaśniają i uszczegóławiają. Zbitki postawy mają za zadanie sygnalizowanie postawy autora wobec opisywanego zagadnienia za pomocą modalności epistemicznej, a także wyrażania obligacji, możliwości, potrzeby, intencji itd.

Hyland (2008a) natomiast podzielił funkcje ZL na takie, które dotyczą opisu badań (*research-oriented*, dalej RO), odnoszące się do organizacji tekstu (*text-oriented*, dalej TO) oraz zbitki dotyczące uczestników dyskursu, tj. autora lub czytelnika tekstu (*participant-oriented*, dalej PO). Zbitki RO dotyczą opisu czasu, miejsca, procedur, ilości, tematu badań. Organizacja tekstu może być zasygnalizowana przez zbitki TO tworzące kontrast (np. *on the other hand*), podkreślające skutek, strukturę tekstu lub ograniczenia w badaniach (np. *in the case of*). Zbitki PO (zorientowane na autora lub czytelnika) mogą dotyczyć wyrażania ewaluacji przez autora (podobnie jak zbitki postawy u Bibera i in., 2004) lub mogą zwracać się bezpośrednio do czytelnika (*as can be seen*).

---

<sup>2</sup> Więcej na temat statystyczny parametr MI, ang. mutual information zob. Kenneth Church i Patrick Hanks 1990, którzy szczegółowo omawiają algorytm obliczający MI.

## **Wczesniejsze badania zbitek językowych i dyskursu medycznego**

Zbitki leksykalne stanowiły przedmiot badań wielu autorów, którzy analizowali rozmaite typy dyskursu. Zbitki leksykalne badano do tej pory w tekstach akademickich (Biber i in. 2004, Hyland 2008a, Hyland i Tsy 2009, Simpson-Vlach i Ellis 2010, Schnur 2014) oraz w tekstach pisanych przez nierodzimych użytkowników języka angielskiego, w kontekście m.in. języka hiszpańskiego, chińskiego, szwedzkiego, czeskiego, koreańskiego, malajskiego, perskiego (Cortes 2006, Neely i Cortes 2009, Kim 2009, Wei i Lei 2011, Ädel i Erman 2012, Dontcheva-Navratilova 2012, Schnur 2014, Kashiha i Chan 2015, Nam i in. 2016, Pan i in. 2016, Esfandiari i Barbary 2017). Oprócz dyskursu akademickiego, ZL badane były również w innych typach tekstów, na przykład w dyskursie politycznym (Partington i Morley 2004), w dokumentach EU (Jablunkai 2010), w tekstach z zakresu telekomunikacji (Pan i in. 2016), w dyskursie prawnym (Goźdź-Roszkowski 2011, Breeze 2013, Kopaczyk 2013) oraz w dyskursie farmaceutycznym (Grabowski 2015, Novais Mazza 2015). Zbitki leksykalne były też analizowane porównawczo w odniesieniu do tekstów reprezentujących dyskursy różnych dziedzin, np. dyskurs historyczny analizowano na tle dyskursu z zakresu biologii (Cortes 2004), dyskurs z zakresu biologii porównywany był też z dyskursem biznesu, inżynierii elektrycznej i językoznawstwa stosowanego (Hyland 2008a) oraz z dyskursem biochemicznym (Salazar 2011). Ponadto, badano szczegółowo zbitki leksykalne w tekstach medycznych (Marco 2000, Jalali i Moini 2014, Jalali i in. 2015). Zbitki językowe w tekstach medycznych to jednak temat badawczy stosunkowo rzadko podejmowany i – o ile mi wiadomo – nie opublikowano do tej pory badań na temat zbitek leksykalnych w artykułach o tematyce medycznej wydawanych w polskich czasopismach naukowych.

Ponieważ język angielski stał się językiem naukowych publikacji medycznych, rośnie liczba fachowych czasopism naukowych poświęconych medycynie oraz samych tekstów medycznych, których tylko w ciągu dwudziestu lat, od roku 1980 do 2000, przybyło około trzech milionów (Gotti i Salager-Meyer 2016:2). Taki ogromny przyrost tekstów medycznych stymuluje badania na temat dyskursu medycznego. Do tej pory opisane zostały takie aspekty dyskursu medycznego, jak np. użycie czasowników modalnych, przymiotników, okresów warunkowych, kolokacji, nauczania języka medycznego itd. (Nwogu 1997, Ferguson 2001, Salager-Meyer i Gotti 2006, Chen i Ge 2007,

Li i Ge 2009, Fryer 2012, Jirapanakorn 2012, Yang i in. 2015, Gotti i Salager-Meyer 2016, Bączkowska w tym tomie), jednak w porównaniu z badaniami na temat innych języków specjalistycznych, tzw. język angielski dla celów medycznych (ang. English for Medical Purposes; EMP) jest stosunkowo rzadko podejmowany.

Dla dydaktyków EMP interesujący może być fakt, że na podstawie korpusów tekstów medycznych powstała lista słów najczęściej używanych w dyskursie medycznym (w medycznych czasopismach naukowych) opracowana przez Wang i in. (2008), która wzorowała się na innej liście – Academic Word List (Coxhead 2000), tj. liście słów najczęściej używanych w dyskursie akademickim w angielskim języku pisanym. Interesujące jest, że Ken Hyland i Polly Tse (2009), badając zbitki leksykalne w kilku dyscyplinach, doszli do wniosku, że listy słów powinny być tworzone dla każdej dyscypliny oddzielnie, bowiem nie sposób ustalić, jakie są podstawowe słowa w specjalistycznym dyskursie akademickim, które odzwierciedlałyby język typowy dla wielu dyscyplin. Z tym wnioskiem nie zgadzają się jednak Simpson-Vlach i Ellis (2010: 509), bowiem udało im się utworzyć listę zawierającą właśnie takie podstawowe słownictwo (podstawową transdyscyplinarną listę fraz) obowiązujące różne dyscypliny specjalistyczne.

## **Dane korpusowe i metodologia badań**

Poniższa analiza przedstawia zbitki leksykalne wyekscerpowane z korpusu dostępnego w systemie Sketch Engine (SK), który składa się z tekstów publikowanych w czasopismach naukowych o wolnym dostępie (korpus DOAJ). Korpus DOAJ zawiera prawie 660 tysięcy artykułów, liczy 2,6 miliarda słów i obejmuje artykuły z wielu dziedzin nauki, w tym z różnych działów medycyny, nauk ścisłych, nauk humanistycznych i społecznych. Korpus ten zawiera metadane, które pozwalają na dokonanie wyboru tekstów do analizy sprofilowanej według np. roku, kraju i języka publikacji czy dziedziny nauki, a tym samym na utworzenie sprofilowanego podkorpusu.

Podzbiór danych oznaczony w systemie SK jako zbiór artykułów o tematyce onkologicznej stanowił podstawę do skompilowania korpusu artykułów o tematyce medycznej opublikowanych w czasopismach polskich (dalej JMP). Korpus JMP liczy ok. 22 mln słów. Zdecydowana większość artykułów w korpusie JMP pochodzi z czasopisma naukowego poświęconego dermatolo-

gii. Po analizie autorstwa tekstów należy przyjąć, że pewna część artykułów zawartych w korpusie ilustruje teksty napisane przez autorów, dla których język angielski oraz język polski są językami obcymi. Jest jednak też znaczna grupa autorów w korpusie JMP, którzy są rodzimymi użytkownikami języka polskiego (oceniając po nazwiskach autorów oraz ich afiliacji). Dla celów porównawczych dane z korpusu JMP zestawiono z danymi uzyskanymi w korpusie naukowych tekstów medycznych stanowiących podkorpus *British National Corpus* (dalej BNC). Narodowy Korpus Brytyjski jest korpusem referencyjnym, który reprezentuje współczesny język angielski, dlatego podkorpus tekstów medycznych, wydzielony z BNC (dalej BNC\_ac\_med) w oparciu o kryteria klasyfikacji opracowane przez Davida Lee (2001), będzie stanowił punkt odniesienia dla badań opisanych poniżej, tj. będzie również funkcjonował jako korpus referencyjny. Jeśli dane w tych dwóch korpusach będą się znacząco różnić, oznaczałoby to, że teksty naukowe publikowane w czasopiśmie wydawanych w Polsce mogą sugerować niższe kompetencje autorów artykułów w zakresie znajomości stylistyki akademickiego języka angielskiego w stosunku do autorów tekstów naukowych publikowanych w Wielkiej Brytanii. Podkorpus tekstów medycznych w BNC liczy ok. 1,4 mln słów, co stanowi 1,45% słów w całym BNC. Artykuły dostępne w BNC\_ac\_med pochodzą z naukowych czasopism medycznych z dziedziny m.in.: gastroenterologii i hepatologii, pielęgniarstwa klinicznego, diabetologii i kardiologii. Zdecydowana większość tekstów opublikowano w prestiżowym czasopiśmie medycznym „British Medical Journal”.

Zbitki leksykalne analizowane poniżej zostały wygenerowane w oparciu o parametr frekwencyjny i wybrane z listy zbitek przy zastosowaniu kilku kryteriów wykluczających. Po pierwsze, za częstość wystąpienia ZL przyjęto 25 zbitek na milion słów w korpusie JMP, co po standaryzacji tekstów wyłoniło 42 spośród 1000 ZL automatycznie wygenerowanych przez oprogramowanie SK. Ponieważ korpusy użyte w badaniu różnią się wielkością, zastosowano dynamiczne podejście do określania dolnego progu minimalnej liczby ZL (Biber i Bibieri 2007, Chen i Baker 2014:854), w związku z czym minimalną liczbę zbitek dla korpusu referencyjnego określono na ok. 29 okurencji na milion (tj. po standaryzacji). Po drugie, wyeliminowano przykłady zbitek, które zawierały specjalistyczne słownictwo medyczne (np. *of platelet activating factor, the lower oesophageal sphincter*) lub były zależne od kontekstu (*in the United Kingdom*), zawierały fragmenty nienapisane przez autora



(informacje od wydawcy dotyczące czasopisma, np. o otwartym dostępie tekstów i prawach autorskich) lub podawały dodatkowe informacje odautorskie (np. afiliację autora). Zbitki zależne od kontekstu, ale wspólne dla obu korpusów, były brane pod uwagę (np. *in the control group*). Po trzecie, badanie zastosowano dla czterowyrazowych zbitek leksykalnych, które są uważane za frazy bardziej miarodajne niż np. zbitki trójwyrazowe czy pięciowyrazowe (Chen i Baker 2010:32, Goźdz-Roszkowski 2011:110). Po czwarte, podobnie jak w badaniach Cortes (2004) oraz Chena i Bakera (2010:32), minimalną liczbę tekstów, w których wystąpiła dana zbitka (parametr dyspersji), określono na 5.

Należy zauważyć, że niektóre zbitki mogą rodzić wątpliwości i, tym samym, różnice w klasyfikacji. Na przykład te same zbitki opisane przez Bibera i in. (2004) oraz Simpson-Vlach i Ellisa (2008) nie zawsze są skategoryzowane w taki sam sposób. Zbitka, *as a result of* jest zaklasyfikowana jako dyskursywna (konkretnie jako podtyp: rozwinięcie tematu) przez Simpson-Vlach i Ellis (2008), ale jako zbitka referencyjna przez Bibera i in. (2004). Podobnie zbitka *the fact that the* pełni funkcję referencyjną (określanie atrybutów) u Simpson-Vlach i Ellis (2008), jednak w propozycji Bibera i in. (2004) spełnia funkcję wyrażania postawy (epistemicznej, bezosobowej). Zbitki leksykalne mogą przyjmować różne funkcje w zależności od kontekstu, np. *the other side of* może występować jako określenia miejsca (zbitka referencyjna) lub jako organizator dyskursu (Cortes 1996:140). Kategoryzacja ZL w niniejszym artykule oparta jest na podziale przyjętym przez Bibera i in. (2004). Zbitki leksykalne, które wzbudzały wątpliwość, były analizowane w rozszerzonym kontekście badanych korpusów (konkordancje, zdania lub akapity).

W tabeli 1 przedstawiono listy ZL wyekscerpowane z korpusu JMP oraz BNC\_ac\_med. Podano frekwencję zbitek (F) oraz ich liczbę po standaryzacji na milion (S). Zbitki, które występują w obydwu korpusach, zaznaczono wytłuszczeniem.

Tabela 1. Frekwencja zbitek leksykalnych w korpusie JMP i BNC\_ac\_med

<i>Korpus JMP</i>	<i>F</i>	<i>S</i>	<i>BNC ac_med</i>	<i>F</i>	<i>S</i>
<b>on the basis of</b> („na podstawie”)	2381	108	<b>at the time of</b> („w czasie”)	163	116
in the case of („w przypadku”)	2194	98	<b>as a result of</b> („w rezultacie”)	129	92
<b>as well as the</b> („jak również”)	1844	84	in the presence of („w obecności”)	121	86
<b>on the other hand</b> („z drugiej strony”)	1782	81	<b>on the basis of</b> („na podstawie”)	118	84
<b>as a result of</b> („w rezultacie”)	1438	65	has been shown to („wykazało”)	109	78
in the form of („w formie”)	1238	56	the end of the („koniec” + rzeczownik)	107	76
is one of the („jest jednym z”)	1219	55	in the number of („w liczbie”)	84	60
the basis of the („podstawa” + rzeczownik)	934	43	in the absence of („wobec braku”)	84	60
in the course of („w toku”)	926	42	was no significant difference („nie było znaczącej różnicy”)	82	59
<b>at the same time</b> („w tym samym czasie”)	827	38	has been shown to („wykazało, że”)	76	54
with the use of („z wykorzystaniem”)	781	36	between the two groups („między tymi dwiema grupami”)	<b>75</b>	54
in the area of („w obszarze”)	740	34	<b>on the other hand</b> („z drugiej strony”)	74	53
one of the most („jeden z najbardziej”)	726	33	at the end of („pod koniec”)	73	52
the fact that the („fakt, że”)	723	33	there was no significant („nie było znaczącej”)	70	50
the aim of this („celem tego”)	698	32	presence or absence of („w obecności lub wobec braku”)	64	46
the aim of the („celem” + rzeczownik)	698	32	is likely to be („jest prawdopodobne, że będzie”)	63	45
<b>in the treatment of</b> („w leczeniu”)	680	31	no significant difference in („brak znaczącej różnicy”)	62	44
in accordance with („w zgodzie z”)	676	31	have been shown to („wykazały”)	60	43
it is possible to („możliwe jest”)	653	30	was approved by the („było zaakceptowane przez”)	58	41
the result of the („wynik” + rzeczownik)	639	29	has been suggested that („zasugerowano, że”)	<b>58</b>	41
as well as in („jak również w”)	628	29	were no significant differences („nie było znaczących różnic”)	57	41
on the other hand („z drugiej strony”)	626	29	studies have shown that („badania pokazały, że”)	56	40
of this study was („tego badania było”)	618	28	<b>in the treatment of</b> („w traktowaniu”)	54	39
the result of the („wynik” + rzeczownik)	606	28	the presence of a („obecność” + rzeczownik)	53	38

it is necessary to („jest konieczne” + czasownik)	604	28	in the present study („w niniejszym studium”)	53	38
the end of the („pod koniec” + rzeczownik)	602	27	in the control group („w grupie kontrolnej”)	53	38
on the basis of („na podstawie” + rzeczownik)	602	27	in the management of („w prowadzeniu”)	51	36
the surface of the („na powierzchni” + rzeczownik)	592	27	there was no significant („nie było znaczącej”)	50	36
conducted according to the („przeprowadzone zgodnie z”)	590	27	the presence or absence („obecność lub brak”)	50	36
the case of the („przypadek” + rzeczownik)	588	27	in the development of („w rozwoju”)	50	36
an important role in („ważną rolę w”)	588	27	to the development of („do rozwoju”)	49	35
was conducted according to („został przeprowadzony zgodnie z”)	585	27	not significantly different from („nieznaczaco różny od”)	<b>48</b>	34
to the fact that („fakt, że”)	585	27	it is possible that („możliwe, że”)	48	34
in the process of („w procesie” + rzeczownik)	580	26	there were no significant („nie było znaczących”)	47	34
the examination of the („badanie” + rzeczownik)	577	26	in the patients with („u pacjentów z”)	46	33
this study was to („tego badania było” + czasownik)	569	26	<b>as well as the</b> („jak również”)	46	33
examination of the patient („badanie pacjenta”)	566	26	of this study was („tego badania było”)	44	31
of the patient was („pacjenta było”)	563	26	were not significantly different („nie były znacząco różne”)	42	30
the patient was conducted („pacjent/a było przeprowadzone”)	560	26	the time of the („czas” + rzeczownik)	42	30
patient was conducted according („pacjent/a było przeprowadzone zgodnie”)	560	26	are shown in table („jak pokazano w tabeli”)	41	29
reported in the literature („opisane w literaturze”)	555	25	a result of the („wynik” + rzeczownik)	41	29
in the context of („w kontekście”)	545	25	at a rate of („w tempie”)	40	29

## Analiza funkcji zbitek leksykalnych

Już pobieżna analiza ZL w obydwu korpusach wykazuje, że użycie zbitek jest porównywalne, tj. najmniej jest zbitek wyrażających postawę, a najwięcej zbitek referencyjnych. Różnica między korpusami jest widoczna w częstotliwości użycia zbitek dyskursywnych, których jest nieznacznie mniej w korpusie referencyjnym oraz w nieznacznie częstszym użyciu zbitek postawy w korpusie referencyjnym. Różnice te nie są jednak statystycznie znaczące, jak wynika

z badania istotności *log likelihood* (w tabeli 2 poniżej oznaczonej jako LL), przy  $p < 0,05$  i wartości krytycznej 3,84.

Tabela 2. Frekwencja i istotność statystyczna zbitok leksykalnych w korpusie JMP i BNC\_ac\_med

ZB	JMP	BNC_ac_me	LL
Dyskursywne	13 (31%)	10 (23,8%)	0,18
Referencyjne	24 (57%)	27 (64,3%)	0,39
Postawy	5 (11,9%)	5 (11,9%)	0

Przykładowe zbitki wyrażające postawę to *the fact that the* („fakt, że”), *it is possible to* („możliwe jest” + czasownik), *it is necessary to* („konieczne jest” + czasownik), *are taken into account* („są brane pod uwagę”), *has been shown to* („pokazano, że”), itp., natomiast zbitki organizujące dyskurs to przykładowo: *the aim of this* („celem tego”), *the result of the* („rezultat” + rzeczownik), *this study was to* („tego studium było”), *as well as in* („jak również w”), *as a result of* („w wyniku”), *on the other hand* („z drugiej strony”) itp. Zbitki ilustrujące referencję to na przykład: *at the same time*, *in the area of*, *is one of the*, *on the basis of*, *in the context of* itp.

Wśród pierwszych dziesięciu ZL, tj. tych o najwyższej frekwencji, są następujące zbitki: *on the basis of* („na podstawie”), *in the case of* („w przypadku”), *as well as the* („jak również”), *on the other hand* („z drugiej strony”), *as a result of* („w wyniku”), *in the form of* („w formie”), *is one of the* („jest jednym z”), *the basis of the* („podstawa” czegoś), *in the course of* („w toku”) oraz *at the same time* („w tym samym czasie”). Dominująca funkcja w korpusie JMP oraz BNC\_ac\_med w pierwszych dziesięciu ZL to funkcja referencyjna (odpowiednio 7 i 6 przypadków), podczas gdy zbitki postawy w ogóle nie występują w JMP i tylko raz w BNC. Nie jest to zaskakująca obserwacja, bowiem zbitki wyrażające postawy oraz organizujące dyskurs są bardziej typowe dla akademickiego dyskursu mówionego niż dla pisanego (Biber 2006, Cortes 2004, Hyland 2008b). W porównaniu do korpusu referencyjnego, pięć ZL spośród dziesięciu najczęściej występujących powtarzają się w korpusie JMP, są to (wytłuszczone w tabeli 1): *on the basis of*, *as a result of*, *on the other hand*, *at the same time* i *as well as the*. Cztery pierwsze to zbitki referencyjne, ostatnia to zbitka dyskursywna. W podkorpusie BNC jest większa różnorodność funkcji zbitok leksykalnych w porównaniu z korpusem JMP.

Analiza korpusu JMP częściowo potwierdza badania Hylanda (2008a: 396-397) dotyczące zbitok leksykalnych w ogólnym dyskursie akademickim,

gdzie ZB występujących w funkcji referencyjnej jest również najwięcej. Nie potwierdza się natomiast obserwacja Hylanda, że w prozie akademickiej jest niewiele zbitek dyskursywnych (nawet mniej niż wyrażających postawę).

## **Wnioski**

Z analizy przedstawionej w niniejszym artykule wynika, że w badanym korpusie tekstów: (i) najwięcej występuje zbitki leksykalnych pełniących funkcję referencyjną; (ii) najrzadziej używane są zbitki leksykalne wyrażające ocenę lub stosunek autora tekstu do opisywanego zagadnienia; (iii) w porównaniu do korpusu referencyjnego, w tekstach publikowanych w polskich czasopismach zbitki dyskursywnych jest nieznacznie więcej, a zbitki postawy jest taka sama liczba. Uogólniając można stwierdzić, że (i) w badanych tekstach medycznych widoczna jest tendencja do nadreprezentacji zbitek referencyjnych i niedoreprezentacji zbitek dyskursywnych i – w szczególności – zbitki postawy; (ii) w porównaniu z tekstami publikowanymi w czasopismach brytyjskich, w tekstach publikowanych w polskich czasopismach naukowych nie zauważono znaczących różnic w użyciu zbitek leksykalnych pod względem funkcji, którą pełnią w tekście.

## **Bibliografia**

- Ädel Annelie, Erman Britt (2012), *Recurrent word combinations in academic writing by native and non-native speakers of English: a lexical bundles approach*, „English for Specific Purposes”, nr 31, s. 81-92.
- Altenberg Bengt (1991), *The London-Lund Corpus: research and applications*, w: „Using Corpora” (Proceedings of the Seventh Annual Conference of the University of Waterloo Centre for the New OED and Text Research), s. 71-83.
- Altenberg Bengt (1998), *On the phraseology of spoken English: the evidence of recurrent word combinations*, w: A.P. Cowie (red.), *Phraseology: Theory, Analysis and Applications. Oxford Studies in Lexicography and Lexicology*, Clarendon Press, s. 101-22.
- Banerjee Satanjeev, Pedersen Ted (2003), *The design, implementation, and use of Ngram Statistics Package*, w: A.F. Gelbukh (red.), *The Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics, February 17-21, 2003*, Mexico City.
- Bączkowska Anna [w tym tomie], *Korpusy językowe w języku specjalistycznym medycznym jako narzędzie wspomagające przekład i nauczanie języka medycznego*.
- Bączkowska Anna [w druku], *Zbitki leksykalne w dyskursie akademickim*.

- Biber Douglas (1996), *University Language: A corpus-based study of spoken and written register*, John Benjamins.
- Biber Douglas, Johansson Stig, Leech Geoffrey, Conrad Susan, Finegan Edward (1999), *Longman Grammar of Spoken and Written British*, Pearson Education.
- 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.
- Biber Douglas, Barbieri Federica (2007), *Lexical bundles in university spoken and written register*, *„English for Specific Purposes”*, nr 26, s. 263-86.
- Breeze Ruth (2013), *Lexical bundles cross four legal genres*, *„International Journal of Corpus Linguistics”*, nr 18/2, s. 229-253.
- Chen Qi, Ge Guang-chun (2007), *A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs)*, *„English for Specific Purposes”*, nr 26, s. 502-514.
- Chen You-hua, Baker Paul (2010), *Lexical bundles in L1 and L2 academic writing*, *„Language Learning and Technology”*, nr 14/2, s. 30-49.
- Chlebda Wojciech (1991), *Wprowadzenie do frazeologii nadawcy*, Wydawnictwo WSP w Opolu.
- Church Kenneth W., Hanks Patrick (1990), *Word association norms, mutual information and lexicography*. *„Computational Linguistics”*, nr 16/1, s. 22-29.
- Cortes Vivian (1996), *Lexical bundles in freshmen composition*, w: R. Reppen, S.M. Fitzmaurice, D. Biber (red.), *Using Corpora to Explore Linguistic Variation*, John Benjamins.
- Cortes Vivian (1996), *Lexical bundles in freshman composition*, w: R. Reppen, S.M. Fitzmaurice i D. Biber (red.), *Using Corpora to Explore Linguistic Variation*, John Benjamins, s. 131-146.
- Cortes Vivian (2004), *Lexical bundles in published and student disciplinary writing: examples from history and biology*, *„English for Specific Purposes”*, nr 23/4, s. 397-423.
- Cortes Vivian (2006), *Teaching lexical bundles in the disciplines: an example from a writing intensive history class*, *„Linguistics and Education”*, nr 17/4, s. 391-406.
- Cortes Vivian (2004), *Lexical bundles in published and student disciplinary writing: examples from history and biology*, *„English for Specific Purposes”*, nr 23, s. 397-423.
- Cortes Vivian (2006), *Teaching lexical bundles in the disciplines: An example from a writing intensive history class*, *„Linguistics and Education”*, nr 17, s. 391-406.
- Coxhead Averil (2000), *A new Academic Word List*, *„TESOL Quarterly”* nr 34, s. 213-238.
- Dontcheva-Navratilova Olga (2012), *Lexical bundles in academic texts by non-native speakers*, *„Brno Studies in English”*, nr 38/2, s. 37-58.

- Esfandiary Rajab, Barbary Fatima (2017), *A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research*, „Journal of English for Academic Purposes”, nr 29, s. 21-42.
- Ferguson Gibson (2001), *If you pop over there: A corpus-based study of conditionals in medical discourse*, „English for Specific Purposes”, nr 20(1), s. 61–82.
- Firth John (1957), *A Synopsis of Linguistic Theory 1930-1955*, „Studies in Linguistic Analysis”, Wydawnictwo Uniwersytetu Oksfordzkiego, s. 1-32.
- Flecher William H. (2006), *Concordancing the Web: promise and problems, tools and techniques*, w: M. Hundt, N. Nesselhauf, C. Biewer (red.), *Corpus Linguistics and the Web*, Brill, s. 25-45.
- Flowerdew Lynn (2012), *Corpora and Language Education*, Basingstoke, Palgrave.
- Fryer Daniel Lees (2012), *Analysis of the generic discourse features of the English-language medical research article: A systemic-functional approach*, „Functions of Language”, nr 19(1), s. 5-37.
- Geeraerts Dirk, Cuyckens Hubert (red.) (2007), *The Oxford Handbook of Cognitive Linguistics*, Wydawnictwo Uniwersytetu Oksfordzkiego.
- Goldberg Adele E. (2006), *Constructions at Work*, Oxford University Press.
- Gotti Maurizio, Salager-Meyer Françoise (2016), *Teaching medical discourse in higher education: an introduction*, „Language Learning in Higher Education”, nr 6/1, s. 1-13.
- Goźdz-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”, nr 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.
- Halliday Michael A.K. (1991), *Corpus Studies and Probabilistic Grammars*, w: K. Aijmer, B. Altenberg (red.), *Advances in Corpus Linguistics*, John Benjamins, s. 30-43.
- Hanks Patrick (2004), *Corpus pattern analysis*, w: *Proceedings of Euralex 2004*, Lorient, Francja.
- Harris Zellig (1954), *Distributional Structure*, „Word”, nr 10(2-3), s. 146-164.
- Hoey Michael (2004), *Textual colligation: a special kind of lexical priming*, w: K. Aijmer, B. Altenberg (red.), *Advances in Corpus Linguistics*, John Benjamins, s. 171-194.
- Hyland Ken (2008a), *“As can be seen”: lexical bundles and disciplinary variation*, „English for Specific Purposes”, nr 27, s. 4-11.
- Hyland Ken (2008b), *Academic clusters: text patterning in published and postgraduate writing*, „International Journal of Applied Linguistics”, nr 18/1, s. 41-62.
- Hyland Ken, Tse Polly (2009), *Academic lexis and disciplinary practice: corpus evidence for specificity*. „International Journal of English Studies”, nr 9/2, s. 111-129.

- Jablonkai Reka (2010), *English „in the context of” European integration: a corpus-driven analysis of lexical bundles in EU documents*, „English for Specific Purposes”, nr 29, s. 253-267.
- Jirapanakorn Narumeol (2012), *How Doctors Report: A Corpus-based Contrastive Analysis of Reporting Verbs in Research Article Introductions Published in International and Thai Medical Journals*, „The Bangkok Medical Journal”, nr 4, s. 39-46.
- Jalali Zahra Sadat, Moini Mohammad Raouf (2014), *Structure of lexical bundles in introduction section of medical research articles*, “Procedia – Social and Behavioural Sciences”, nr 98, s. 719-726.
- Jalali Zahra Sadat, Moini Mohammad Raouf, Arani Mohamad Alaei (2015), *Structural and Functional Analysis of Lexical Bundles in Medical Research Articles: A Corpus-Based Study*, „International Journal of Information Science and Management”, nr 13/1, s. 51-69.
- Kashiha Hadi, Chan Swee Heng, (2015), *A little bit about: differences in native and non-native speakers’ use of formulaic language*, „Australian Journal of Linguistics”, nr 35/4, s. 297-310.
- Kim YouJin (2009), *Korean lexical bundles in conversation and academic texts*, „Corpora”, nr 4/2, s. 135-165.
- Kjellmer Göran (1987), *Aspects of English collocations*, w: W. Meijs (red.), *Corpus Linguistics and Beyond* (Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora), Rodopi, s. 133-40.
- Kopaczyk Joanna (2013), *Legal Language of Scottish Burghs: Standardization and Lexical Bundles*, Wydawnictwo Uniwersytetu Oksfordzkiego.
- Marco Maria Jose Luzon (2000), *Collocational frameworks in medical research papers: a genre-based study*, „English for Specific Purposes”, nr 19, s. 63–86.
- Nam Kilim, Song Hyun-ju, Choi Jun [2016], *A morpheme-based analysis of lexical bundles in Korean: an interface between corpus-driven approach and lexicography*, „Lexicography”, nr 3/1, s. 39-62.
- Neely Elizabeth, Cortes Vivian (2009), *A little bit about: analysing and teaching lexical bundles in academic lectures*, „Language Value”, nr 1/1, s. 17-38.
- Novais Mazza, Luciene (2015), *Processamento linguístico-computacional de pacotes lexicais: um estudo de corpus na área de Regulamentação Farmacêutica*, „Calidoscópico”, nr 13/3, s. 424-439.
- Langacker Ronald (1987), *Foundations of Cognitive Grammar*, Stanford University Press.
- Lee David (2001), *Genres, register, text types, Routines and styles: clarifying the concepts and navigating a path through the BNC jungle*, „Language Learning and Technology”, nr 5/3, s. 37-72.



- Li Juan-Li, Ge Guang-Chun (2009), *Genre analysis: Structural and linguistic evolution of the English medium medical research article (1985–2004)*, „English for Specific Purposes”, nr 28(2), s. 93-104.
- Nwogu Kevin Ngozi (1997), *The medical research paper: Structure and functions*, „English for Specific Purposes”, nr 16(2), s. 119-138.
- Pahta Paivi (2006), *This is very important: a corpus study of amplifiers in medical writing*, w: F. Salager-Meyer i M. Gotti (red.), *Advanced in Medical Discourse: Oral and Written Contexts*, Peter Lang, s. 357-382.
- Pan Fan, Reppen Randi, Biber Douglas (2016), *Comparing patterns of L1 versus L2 English academic professional: lexical bundles in telecommunications research*, „Journal of English for Academic Purposes”, nr 21, s. 60-71.
- Partington Alan, Morley John (2004), *From frequency to ideology: investigating word and cluster/bundle frequency in political debate*, w: B. Lewandowska-Tomaszczyk (red.), *Practical Applications in Language and Computers*, Peter Lang, s. 179-192.
- Salager-Meyer Froncoise, Gotti Maurizio (red.) (2006), *Advanced in Medical Discourse: Oral and Written Contexts*. Bern: Peter Lang.
- Salazar Danica (2011), *Lexical Bundles in Scientific English: A corpus-based study of native and non-native writing*. Nieopublikowana praca doktorska, Uniwersytet w Barcelonie. Dostęp: <http://diposit.ub.edu/dspace/bitstream/2445/35023/2/DJLS DISSERTATION.pdf> (5.03.2018).
- Schnur Erin (2014), *Phraseological signaling of discourse organization in academic lectures: a comparison of lexical bundles in authentic lectures and EAP listening materials*, „Yearbook of Phraseology”, nr 5/1, s. 95-122.
- Scott Michael, Tribble Christopher (1997), *Textual Patterns: Keywords and Corpus Analysis in Language Education*, Amsterdam: 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.
- Sinclair John (1991), *Corpus, Concordance, Collocation*, Oxford University Press.
- Skorupka Stanisław (1952), *Z zagadnień frazeologii*, „Sprawozdania Komisji Językowej TNW”, IV, Wydawnictwo Towarzystwa Naukowego Warszawskiego.
- Stephanowitsch Anatol, Gries Stefan Th. (2003), *Collostructions: investigating the interaction of words and constructions*, „International Journal of Corpus Linguistics”, nr 8(2), s. 209-243.
- Stubbs Michael (1996), *Text and Corpus Linguistics*, Blackwell.
- Tognini-Bonelli Elena (2001), *Corpus Linguistics at Work*, John Benjamins.
- 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.
- Wei Yaoyu, Lei Lei (2011), *Lexical bundles in the academic writing of advanced Chinese EFL learners*, „RELC Journal”, nr 42/2, s. 155-166.
- West Michael (1953), *A General Service List of English Words*, Longman.

Winogradow Wiktor Wladimirowicz (1947), *Ob. Osnownych tipach frazeologiczeskich jedinic*, „Trudy komisji po istori AN SSSR”, nr 3.

Yang An, Zheng Shu-yuan, Yuan, Ge, Guang-chun (2015), *Epistemic modality in English-medium medical research articles: A systemic functional perspective*, „English for Specific Purposes”, nr 38, s. 1–10.

## Functional analysis of lexical bundles in medical discourse

**Summary:** The paper describes lexical bundles (Biber 1996) occurring in research medical texts. Two corpora of medical texts are used to analyse lexical bundles both qualitatively and quantitatively: a corpus of articles published in Polish journals and a corpus of texts published in British journals (largely in the *British Medical Journal*). The investigation reveals that referential lexical bundles are prevalent in both corpora. No significant differences can be observed between the two corpora, yet there is a tendency to use more discursive lexical bundles in papers published in Polish journals.

**Keywords:** *lexical bundles, medical discourse, academic discourse, functional analysis*

**Słowa kluczowe:** *zbitki leksykalne, dyskurs medyczny, dyskurs akademicki, analiza funkcji*