

## Kolokacje w tekstach prawnych – problemy proceduralne

Collocations in legal texts – procedural issues

Małgorzata Gębka-Wolak

Wydział Filologiczny, Uniwersytet Mikołaja Kopernika w Toruniu,  
ul. Fosa Staromiejska 3, 87-100 Toruń, Polska;  
e-mail: mge@umk.pl

Andrzej Moroz

Wydział Filologiczny, Uniwersytet Mikołaja Kopernika w Toruniu,  
ul. Fosa Staromiejska 3, 87-100 Toruń, Polska;  
e-mail: amoroz@umk.pl

### Abstrakt

Artykuł wpisuje się w krąg badań, których celem jest przedstawienie ograniczeń w budowaniu konstrukcji składniowych w tekstach prawnych. Za jeden z przejawów tych ograniczeń uznaje się swoistą łączliwość leksykalną, opisywaną za pomocą pojęcia kolokacji. Kolokacją nazywa się wyrażenie składające się z komponentów, których współwystępowanie jest motywowane leksykalnymi ograniczeniami ich łączliwości. Celem poniższego opracowania jest objaśnienie wstępnej, częściowo zautomatyzowanej procedury rozpoznawania kolokacji w podkorpusie tekstów prawnych z dziedziny prawa karnego.

**Słowa kluczowe:** język specjalistyczny; tekst prawny; kolokacja; analiza automatyczna.

### Abstract

This article is a part of research dedicated to describe restrictions in the building syntactic structures in legal texts. One of the manifestations of these restrictions is a specific lexical connection, that is described by the concept of collocation. Collocation is defined as an expression consisting of components, whose coexistence is motivated by the lexical restrictions of their connectivity. The aim of this study is to explain the initial partially automated procedure undertaken in recognizing collocations in subcorpus of legal texts – from criminal law.

**Keywords:** language for special purposes; legal text; collocation; automatic text analysis.

## O. Założenia

Proponowane niżej obserwacje wpisują się w obszar podjętych niedawno przez autorów tego opracowania badań nad własnościami tekstów prawnych (zob. Gębka-Wolak 2014; Moroz 2014b), czyli tekstów, za pomocą których reguluje się stosunki między członkami danej społeczności, a więc tworzy się prawo. Są to teksty konkretnych aktów prawodawczych, tj. ustaw, rozporządzeń, zarządzeń czy uchwał (Zieliński 1999: 50; 2004: 11, 13). Zakładamy, że jedną z cech, która teksty prawne różni od tekstów wykorzystywanych do innych celów niż tworzenie prawa, jest swoista łączliwość wyrazowa, to jest używanie w tekstach prawnych takich połączeń wyrazowych (kolokacji), które w tekstach nienależących do prawnych nie występują lub występują z wyraźnie mniejszą frekwencją. Zasadniczym celem prowadzonych badań jest ujawnienie specyficznych kolokacji używanych w tekstach prawnych, a co za tym idzie – opisanie ograniczeń w budowaniu konstrukcji składniowych. Nie ulega wątpliwości, że punktem wyjścia tak zakrojonych badań musi być zdefiniowanie procedur wyróżniania kolokacji właściwych dla tekstów prawnych. Jest to zarazem zadanie realizowane przez prezentowany tu zestaw artykułów, w którego skład poza niniejszym tekstem, objaśniającym procedurę wstępnego, częściowo zautomatyzowanego procesu rozpoznawania kolokacji, wchodzi jeszcze dwa kolejne teksty. Ich zadaniem jest zarysowanie zasad przeprowadzania jakościowej analizy ciągów wytypowanych do analizy w wyniku zastosowania procedur wchodzących w skład pierwszego etapu identyfikacji kolokacji. Posłużymy się przy tym dwoma przykładowymi wyrażeniami, a mianowicie strukturami *wykonanie orzeczenia* oraz *czyn zabroniony*.

Z uwagi na domniemanie, że poszczególne gałęzie prawa mogą w zakresie podejmowanych analiz różnić się między sobą i dążąc do zachowania homogeniczności zbioru charakteryzowanych wyrażen, podłożem też badawczych, jak również bezpośrednim przedmiotem opisu uczyniliśmy pięć kodeksów reprezentujących prawo karne, a mianowicie: *Kodeks karny*, *Kodeks postępowania karnego*, *Kodeks karny wykonawczy*, *Kodeks wykroczeń* oraz *Kodeks postępowania w sprawach o wykroczenie*. Ograniczenie takie jest motywowane faktem istnienia ścisłej korelacji między łączliwością wyrażen a aktem normatywnym, w którym są one używane. Jak zauważa bowiem Andrzej Malinowski, „tendencja do wspólnego występowania poszczególnych słów w tekście [...] jest cechą charakterystyczną poszczególnych aktów normatywnych, co wiąże się z odrębnością przedmiotu regulacji tych aktów” (Malinowski 2006: 138).

0.1. Analiza tekstów prawa wiąże się z koniecznością ustosunkowania się do szeregu motywujących ich istnienie dychotomii (m.in. język prawny a ogólny, język prawny a urzędowy czy język prawny, prawniczy a języki okołoprawne, zob. Choduń 2004, Petzel 2011, Zieliński 2011, Moroz 2014a), jak również rozwiązania szeregu problemów teoretycznych, m.in. kwestii zrozumiałości takich tekstów (zob. Gizbert-Studnicki 2009, Jadacka 2010) czy ich zgodności z normą polszczyzny

ogólnej (zob. Choduń 2006, Kącka-Rodak, Wichrowska 2011, Petzel 2011, Gębka-Wolak 2014, Moroz 2014a). Przedyskutowanie wszystkich zagadnień teoretycznych wiążących się z podejmowanym tematem przekracza możliwości krótkiego tekstu. Wydaje się jednak, że jedno z nich jest istotne z punktu widzenia prowadzonych tu rozważań. Otóż uznaje się, że w sytuacji idealnej stosunek formy do treści wyrażenia jest odwzorowywany w relacji 1:1, czyli określony sens jest oddawany przez jedną formę. Prawdą jest jednak to, że takie powiązania w języku nie są zbyt częste, a użytkownik w gruncie rzeczy zmagą się nieustannie z rozbieżnościami zachodzącymi na linii treść – forma<sup>1</sup>. Idą one bądź to w kierunku przypisywania danej formie więcej niż jednej treści, czego efektem jest polisemia, bądź oddawania tej samej treści za pomocą więcej niż jednej formy, co prowadzi do synonimii. Problem napięć pomiędzy poziomami języka – co istotne – dotyka także tekstów prawnych, gdzie dochodzi do zderzenia normy prawnej i jej językowego wykładnika – przepisu. Każdy członek społeczności jest bowiem zobligowany do przestrzegania zasad określonych normami, jednak dostęp do nich uzyskuje poprzez ich językowy korelat. Tym samym im gorszy jest formalny wykładnik prawa, tym trudniej zidentyfikować kompleks zachowań wynikających z normy. Im większa jest rozbieżność sensu wynikającego z przepisu prawnego interpretowanego czysto językowo, zgodnie z zasadami obowiązującymi w komunikacji niespecjalistycznej, i sensu wynikającego z interpretacji uwzględniającej zasady budowania tekstu prawnego, tym większe prawdopodobieństwo, że użytkownik języka niewłaściwie odczyta narzucane mu zasady postępowania. Z takiego stwierdzenia wyłania się pierwsza przesłanka prowadzonych tu rozważań – że zdefiniowanie środków językowych wykorzystywanych w przepisach winno być realizowane z rozwagą, gdyż wpływa to bezpośrednio na sposób odbioru sensu normy prawnej przez odbiorcę (por. Choduń 2009, Łapa 2008).

0.2. Drugi problem dotyczy rozumienia pojęcia kolokacji, które w istniejącej literaturze przedmiotu nie jest jednoznaczne<sup>2</sup>. Jak się wydaje, w procesie interpretacji tego terminu da się wyróżnić dwa podstawowe podejścia<sup>3</sup>. Pierwsze z nich wychodzi od kryterium ilościowego (frekwencyjnego). Kolokacją w tym sensie jest połączenie takich elementów leksykalnych, które w tekstach określonego języka wykazują tendencję do występowania łącznie (są używane częściej niż inne połączenia budowane na tych samych segmentach). Zasób kolokacji danego

<sup>1</sup> Na gruncie współczesnego językoznawstwa powszechnie przyjmuje się założenie o braku jednoznacznego przyporządkowania między formą a treścią znaku. Jednocześnie wskazuje się na to, że zależność między formą a treścią jednostek językowych istnieje, jest ściśle określona i nieprzypadkowa, a co za tym idzie, trzeba ją uwzględniać w opisie (zob. Karolak 1984: 17–18; Zaron 2009: 8; Апресян 2006). Jest natomiast jasne, że opis ten może być prowadzony za pomocą różnych metod i narzędzi, zgodnie z założeniami danego modelu opisu języka i przyjętą siatką pojęć.

<sup>2</sup> Lista definicji kolokacji przywoływanych w literaturze w latach 1957–2008 została zamieszczona w pracy Violety Seretan (2011: 131–132).

<sup>3</sup> Idea wyróżniania dwóch typów definicji kolokacji pochodzi z tekstu Kiry Peshkov (2013: 175–176).

języka lub ich odmiany ustala się tu na podstawie badań statystycznych. Przy drugim rozumieniu kolokacji – motywowanym *de facto* kryterium jakościowym (leksykalno-semantycznym) – jest ona rozumiana jako ograniczone leksykalnie współwystępowanie jednostek wyrazowych (Preshov 2013: 176). Przy tej interpretacji wysoka frekwencja danego połączenia jest jedynie sygnałem – jednak nierozstrzygającym – faktu, że mamy do czynienia z kolokacją, a obrona tezy o kolokacyjnym charakterze danego połączenia jest realizowana z wykorzystaniem narzędzi analitycznych, umożliwiających wskazanie ograniczeń łączliwości.

W prowadzonych w tym miejscu rozważaniach przyjmuje się, że kolokacja jest takim połączeniem wyrażeń, które spełnia oba przytoczone warunki – ilościowy i jakościowy – jednak ich rola w identyfikacji i opisie obiektów analizy jest inna. Otóż cechy frekwencyjne danego połączenia służą wyodrębnieniu zbioru wyrażeń podejrzewanych o bycie kolokacją. Narzędzia stosowane na tym etapie mają charakter zautomatyzowany i ich zadaniem jest oddzielenie połączeń częstych (sygnalizujących pewną tendencję) od nieistotnych. Tak stworzony zbiór winien jednak być poddany ocenie jakościowej, której zadanie polega na wskazaniu leksykalnych i formalnych ograniczeń łączliwości. Ograniczenia te przejawiają się m.in. w taki sposób, że komponenty kolokacji są dobierane z uwzględnieniem postaci leksykalnej przeciwcłonu. Z formalnego punktu widzenia kolokacja jest podrzędną grupą syntaktyczną<sup>4</sup>, którą można zdefiniować morfologicznie, ze względu na klasę gramatyczną reprezentanta i budowę wewnętrzną.

Przyjmujemy jednocześnie, że identyfikacja kolokacji odbywa się kilkuetapowo. W pierwszej kolejności za pomocą zautomatyzowanej procedury wyodrębnia się zbiór wyrażeń, które w tekstach mają tendencję do współwystępowania<sup>5</sup>. Dla uproszczenia wyrażenia te także nazywamy kolokacjami, choć są nimi jedynie w ujęciu ilościowym. Następnie wyodrębniony na podstawie kryterium ilościowego zbiór poddaje się procedurze weryfikacji za pomocą kryteriów formalnych i leksykalno-semantycznych. Pierwszy typ kryteriów (formalne) służy do odrzucenia tzw. kolokacji fałszywych, to jest segmentów, które na linii tekstu ze sobą sąsiadują, ale nie stanowią członu syntaktycznego, jak np. *przepisy niniejszego* w ciągu *przepisy niniejszego kodeksu*, a także *podlega kto* w ciągu *tej samej karze podlega, kto dopuszcza do popełnienia czynu określonego w § 1*.

<sup>4</sup> Grupą syntaktyczną jest każdy fragment tekstu wyodrębniony na dowolnym etapie oddolnej (redukcyjnej) analizy składnikowej (zob. Saloni, Świdziński 2007: 268).

<sup>5</sup> Warto dodać, że wykorzystanie zautomatyzowanych procedur do wstępnej selekcji materiału i wytypowania wyrażeń, które zostaną poddane analizie przy zastosowaniu kryteriów ilościowych, jest metodą stosowaną np. podczas tworzenia kombinatoryczno-objaśniających słowników terminologii specjalistycznej. Na przykład podczas tworzenia słownika terminologii informatycznej (*Dictionnaire fondamental de l'informatique et de l'Internet*) pierwszy etap pracy polegał na automatycznym porównaniu korpusów informatycznego i ogólnego w celu wyszukania jednostek o nietypowej frekwencji, tj. w korpusie informatycznym mających zdecydowanie większą frekwencję niż w korpusie ogólnym. W stosunku do wyszukanej automatycznie listy hipotetycznych terminów specjalistycznych zostały następnie zastosowane kryteria leksykalno-semantyczne (zob. L'Homme 2005: 1123).

Wszystkie zdefiniowane wyżej założenia pozwalają na podjęcie głównego wątku niniejszego tekstu, czyli scharakteryzowanie wieloetapowej procedury identyfikacji struktur kolokacyjnych, co będzie przedmiotem kolejnego punktu.

## 1. Procedura wyboru

Punktem wyjścia do opisu kolokacji winna być procedura identyfikacyjna. Przyjmuje się tu, że jest ona wieloetapowa i opiera się zasadniczo na wskaźnikach frekwencyjnych. Zostanie przeprowadzona za pomocą darmowego analizatora AntConc w wersji 3.4.3. Procedura obejmuje zautomatyzowane wskazanie wyrażań o najwyższej frekwencji tekstowej oraz – na drugim etapie – przypisanie takim wyrażeniom statusu tzw. potwierdzonych kolokatów.

### 1.1. Lista frekwencyjna zlematyzowanych<sup>6</sup> form wyrazowych

Identyfikację kolokacji rozpoczyna tworzenie listy frekwencyjnej zbiorów form występujących w tekstach prawnych. Jest bowiem rzeczą oczywistą, że opis wszystkich możliwych połączeń wyrazowych jest niewykonalny w krótkim czasie, a tym samym warto skoncentrować się na wyrażeniach najczęstszych. Należy przy tym odróżnić frekwencję pojawiających się form wyrazowych od frekwencji zbiorów tych form – leksemów. Uznając, że różne formy fleksyjne realizują w gruncie rzeczy ten sam typ kolokacji (por. *podlega karze* i *podlegał karze*), do badań włącza się zbiory form. Na wstępnym etapie analiz uznano przy tym, że obserwacją zostaną objęte rzeczowniki, których frekwencja tekstowa przekracza 200 wystąpień oraz czasowniki o frekwencji powyżej 50.

Na tym etapie należy brać – niestety – pod uwagę fakt, że automatyczna lematyzacja w gruncie rzeczy wyklucza możliwość mechanicznego rozwiązywania homonimii (podobny problem *de facto* występuje w NKJP (zob. Przepiórkowski i in. 2012: 69–71), gdzie podczas komputerowego przetwarzania tekstu słowa homonimiczne są standardowo kwalifikowane do wszystkich możliwych lematów, co naturalnie fałszuje otrzymane dane frekwencyjne). Z tego też powodu w prowadzonych badaniach przyjęto zasadę, że homonimy takie kwalifikuje się do leksemów zgodnie z największym prawdopodobieństwem wystąpienia (np. forma *kary* zostanie zaliczona do rzeczownika KARA, nie zaś do przymiotnika KARY)<sup>7</sup>.

---

<sup>6</sup> Lematyzacja to termin wykorzystywany na gruncie lingwistyki korpusowej. Jest to wykonane za pomocą narzędzi informatycznych automatyczne oznakowanie leksykalne danego korpusu tekstów, polegające na powiązaniu słów tekstowych (tzw. tokenów) z odpowiednim leksemem. Na zestawionej automatycznie liście frekwencyjnej zbiory słów tekstowych są reprezentowane przez lemat, czyli podstawową (hasłową) formę danego leksemu (Przepiórkowski i in. 2012: 88, 102).

<sup>7</sup> Technicznie rozwiązuje się to przez włączenie do programu AntConc listy prawdopodobnych lematów i reprezentowanych przez nie zbiorów form wyrazowych.

## 1.2. Wybór kolokatów

Kolejny etap procedury identyfikacyjnej prowadzi do wskazania kolokatów tzw. potwierdzonych. W tym celu przygotowuje się listy frekwencyjne wyrazów najczęściej występujących w bezpośrednim lewostronnym i prawostronnym kontekście form wyrazowych należących do jednego leksemu (bada się tym samym łączliwość leksemu nie zaś poszczególnych jego form). Na tym etapie określa się też procentowy udział danej kolokacji w całkowitej liczbie wszystkich kolokacji danego leksemu. Jeżeli wartość ta przekracza 10%, dany kolokat uznaje się za potwierdzony. Wskaźnik dziesięcioprocentowy został ustalony aposteriorycznie przy uwzględnieniu faktu, że w większości wypadków stanowi on wartość graniczną częstych kolokatów danego lematu.

Ten etap analizy narażony jest jednak również na kilka czynników fałszujących. Przede wszystkim uwzględniane kolokaty znajdują się w bezpośrednim kontekście lewo- i prawostronnym, jest natomiast oczywiste, że mogą one wystąpić też w kontekście niebezpośrednim. Problemem jest jednak to, że przy uwzględnieniu takich połączeń proporcje pomiędzy liczbą wszystkich kolokacji a kolokacjami zawierającymi dany kolokat ulegają znacznemu zachwianiu, jak również rośnie prawdopodobieństwo występowania kolokacji tzw. fałszywych. Dodatkowo pewną grupę połączeń mogą stanowić niepodzielne jednostki języka. Ich opis odbiega od charakterystyki konstrukcji, jakimi są kolokacje. Udowodnienie faktu, że dane połączenie jest niepodzielną jednostką, winno być jednak oparte na dokładnych analizach semantyczno-formalnych, czego nie da się przeprowadzić w sposób zautomatyzowany.

## 1.3. Ustalenie cech kolokacji i weryfikacja kolokacji fałszywych

Na podstawie wcześniejszych etapów identyfikacyjnych ustala się kształt kolokacji z wyróżnionym ośrodkiem kolokacji<sup>8</sup> (bazą) i kolokatem. W sytuacji, gdy frekwencja tekstowa danego kolokatu przekracza 10% w stosunku do całkowitej liczby kolokatów danego leksemu, uznaje się go za kolokat potwierdzony, a samą kolokację – za mającą charakter jednostronny (ograniczenie jest nakładane przez bazę na zespół kolokatów). Przykładowo, w badanych tekstach prawnych wyrażenie *wniosek dowódcy* zostało uznane za kolokację jednostronną z ośrodkiem *dowódcy* i kolokatem *wniosek*. Należy przez to rozumieć, że co najmniej 10% wszystkich wystąpień form leksemu *dowódca* łączy się z formą należącą do leksemu *wniosek*. Z kolei dla form leksemu *wniosek* połączenia z formami leksemu *dowódca* nie przekroczyły 10-procentowego progu frekwencji.

---

<sup>8</sup> Ośrodek kolokacji to wyrażenie, dla którego na podstawie badań frekwencyjnych jesteśmy w stanie wskazać taki element leksykalny, zwany kolokatem, z którym jest ono używane wyrażnie częściej niż z innymi elementami leksykalnymi.



Powyższe etapy pozwalają z kolei na wstępne ustalenie cech gramatycznych wyróżnionych kolokacji. W tym celu przygotowuje się listę frekwencyjną klastrów kolokacyjnych<sup>9</sup> wraz z ich pozycją rangową i wyrażoną liczbowo częstością występowania. Działanie takie pozwala wykryć istniejące ograniczenia w budowie kolokacji (m.in. ograniczenia wyboru formy kolokatu czy szyku całej konstrukcji). Celem dodatkowym jest natomiast wyeliminowanie kolokacji tzw. fałszywych, czyli połączeń przypadkowych (występujące obok siebie wyrażenia nie mają ze sobą nic wspólnego, np. *wolności podlega*) oraz przejawów typowych realizacji mechanizmów systemowych (np. kolokat *o* dla gerundium *dokonanie* w połączeniu *o dokonanie*). Opis takich połączeń jest albo niemożliwy, albo zbędny z uwagi na ich typowość.

#### 1.4. Wykrycie kolokacji dwustronnych

Ostatnim elementem procedury identyfikacyjnej jest podjęcie problemu kolokacji tzw. odwróconej. Przyjmuje się w tym wypadku, że bazą kolokacji jest leksem kolokatu etapu pierwszego. Jeżeli frekwencja tekstowa kolokatu takiej bazy przekracza 10%, to kolokat jest uznawany za potwierdzony, a sama kolokacja ma charakter jednostronny ze ściśle określonym ośrodkiem. Jeżeli dodatkowo w kolokacji AB ponaddziesięcioprocentowa frekwencja kolokatu A dla bazy B pokrywa się z ponaddziesięcioprocentową frekwencją kolokatu B dla bazy A, mamy do czynienia z kolokacją dwustronną: nie istnieje tu ośrodek kolokacji, a człony wzajemnie się warunkują. Taka sytuacja występuje w wypadku połączenia *podlega karze*, w którym dla bazy *podlegać* kolokat *kara* stanowi 33,15% wszystkich kolokatów, a przy kolokacji odwróconej kolokat *podlegać* dla bazy *kara* stanowi 15% wszystkich kolokatów.

## 2. Wnioski z analiz frekwencyjnych

Do analiz – jak wspomniano wcześniej – wybrano formy czasownikowe i rzeczownikowe charakteryzujące się wysoką frekwencją tekstową. Realizowały one łącznie 170 leksemów, w tym 130 rzeczownikowych oraz 40 czasownikowych. Przykłady 10 najczęstszych rzeczowników i czasowników wraz z ich frekwencją tekstową w badanym korpusie kodeksów oraz pozycją na liście rangowej zawiera poniższa tabela<sup>10</sup>.

<sup>9</sup> Klastrem kolokacyjnym nazywamy co najmniej dwuelementowy układ wyrażen ustalonych pod względem szyku. Tym samym kolokacja oddaje wszystkie możliwe połączenia ośrodka – bez uwzględniania prepozycyjnego i postpozycyjnego położenia kolokatu – a klastrowy wskazuje już jednoznacznie na fakt, w ilu wystąpieniach dany kolokat lokował się w prepozycji ośrodka kolokacji, w ilu zaś znalazł się w jego postpozycji.

<sup>10</sup> Tworzenie list rangowych jest istotnym elementem badań na strukturą języka prawnego (por. Malinowski 2006, Grzelak 2010).

Tabela 1. Przykłady najczęstszych rzeczowników i czasowników w badanym korpusie kodeksów

RZECZOWNIKI			CZASOWNIKI		
Ranking	Frekwencja	Lemat	Ranking	Frekwencja	Lemat
11	1972	SĄD	18	1229	BYĆ
13	1850	KARA	19	1193	MÓC
16	1328	OSOBA	32	834	PODLEGAĆ
20	1182	POSTĘPOWANIE	45	599	STOSOWAĆ
21	1165	WOLNOŚĆ	66	399	MIEĆ
25	940	ORZECZENIE	78	309	MOŻNA
29	902	SKAZANY	83	288	ZOSTAĆ
31	839	POZBAWIENIE	99	247	NALEŻY
33	802	WYKONANIE	107	228	STANOWIĆ
34	740	PRZESTĘPSTWO	115	208	PRZYSŁUGIWAĆ

Poddane analizie wyrażenia konstituują łącznie 405 kolokacji, z czego 53 (około 13% wszystkich kolokacji) to kolokacje dwustronne, a 352 (około 87%) – jednostronne. W grupie drugiej (jedynie tam istnieje możliwość wskazania ośrodka kolokacji) bazą kolokacyjną bywa – zgodnie z częstością wystąpienia – czasownik (143 wystąpienia), rzeczownik (135 wystąpień), przymiotnik (72 wystąpienia) oraz przyimek i przysłówki (po 1 wystąpieniu). Z kolei kolokatem w takich strukturach bywa rzeczownik (260 kolokacji), przyimek (49 kolokacji), czasownik (37 kolokacji) oraz przymiotnik (5 kolokacji) i przysłówki (1 kolokacja).

Tabela 2. Klasy gramatyczne elementów budujących kolokacje jednostronne

KOLOKACJE JEDNOSTRONNE		
KLASA GRAMATYCZNA	OŚRODEK	KOLOKAT
rzeczownik	135	260
czasownik	143	37
przymiotnik	72	5
przyimek	1	49
przysłówki	1	1

Powyższe dane frekwencyjne pozwalają wysnuć wniosek, że w analizowanych tekstach prawnych dominujący charakter mają kolokacje jednostronne, a w tak wyodrębnionym zbiorze najistotniejszą rolę odgrywają wyrażenia rzeczownikowe (w tym gerundialne), które niezwykle często pełnią funkcję zarówno ośrodka kolokacji, jak i kolokatu, oraz czasownikowe, które są predestynowane do roli bazy kolokacji.



### 3. Specyficzne własności kolokacji

Przeprowadzona procedura pozwoliła na wyodrębnienie struktur uznawanych na podstawie kryterium frekwencyjnego za kolokacje. Jest ona jednak dopiero punktem wyjścia do jakościowej charakterystyki takich obiektów. Jej celem jest z jednej strony wskazanie połączeń specyficznych i oddzielenie ich od struktur typowych dla języka ogólnego oraz – z drugiej strony – opis cech formalnych i znaczeniowych kolokacji prawnych, które docelowo należy uwzględnić w opracowaniach leksykograficznych. Z uwagi jednak na fakt, że analiza szerszej partii materiału w krótkim tekście nie jest możliwa, przedmiotem obserwacji będą jedynie ogólne właściwości dwóch zasadniczych typów kolokacji, tj. kolokacji dwustronnych i jednostronnych. Stanowi to punkt wyjścia do szczegółowej charakterystyki dwóch przykładowych kolokacji, który to wątek zostanie podjęty w dwóch kolejnych artykułach niniejszego cyklu.

#### 3.1. Kolokacje dwustronne

Kolokacje dwustronne są połączeniami dwuelementowymi, w których każdy z członów stanowi dla oponenta potwierdzony kolokat (należy do wyrażen przekraczających 10% kolokatów przeciwczłonu). W wypadku takich struktur uzasadnione jest – z uwagi na silne powiązania wzajemne – rozważenie hipotezy o uznaniu ich za wielosegmentowe jednostki języka, względnie – co bardziej prawdopodobne – za potencjalne terminy języka prawnego. W grupie tej dominują połączenia dwurzeczownikowe (typ A), o następującej charakterystyce gramatycznej: (a) gerundium [N(GER)] + postpozycyjny rzeczownik niegerundialny w dopełniaczu [N(gen)], por. *dokonywanie zakupów, ograniczenie wolności, pozbawianie wolności*, (b) rzeczownik niegerundialny [N] z postpozycyjnym rzeczownikiem gerundialnym w dopełniaczu [N(GER(gen))], por. *akt oskarżenia, zakaz prowadzenia, kara pozbawienia*, (c) połączenia dwóch rzeczowników gerundialnych [N(GER)]+[N(GER)], por. *wykonanie orzeczenia, zatarcie skazania, popełnienie wykroczenia*, (d) połączenie dwóch rzeczowników niegerundialnych [N]+[N], *organ państwa, okres próby, termin x dni*, por.

(1) *W razie skazania na karę **pozbawienia wolności** [...] **zatarcie skazania** następuje z mocy prawa...* KK, 107

(2) *Sąd może orzec **zakaz prowadzenia** określonej działalności gospodarczej...* KK, 41,2

(3) *Zespół powołują, w drodze porozumienia, Prokurator Generalny oraz właściwy **organ państwa**, o którym mowa w § 1.* KPK 589b, 2

Zdecydowanie rzadsze są struktury czasownikowo-nominalne (typ B), ze względu na budowę reprezentujące dwa podtypy: (a) z rzeczownikiem – często gerundialnym – w pozycji niepodmiotowej, por. *uzasadnia przekonanie, podlega karze*, lub (b) z rzeczownikiem w pozycji podmiotowej, por. *ściganie następuje*,

stosuje się. Mało liczne są kolokacje dwustronne będące strukturami dwuczasiownikowymi (typ C), por. *obiecuje udzielić, powinien zawierać*, czasownikowo-przysłówkowymi (typ D), por. *rażąco narusza, stanowi inaczej, dobrowolnie uchyli*, czy rzeczownikowo-przymiotnikowymi (typ E), por. *tymczasowe aresztowanie, postępowanie karne, warunkowe zwolnienie, zakład karny, czyn zabroniony, areszt śledczy, kodeks karny*, np.

(4) *Grzywny nie orzeka się, jeżeli [...] jego stosunki majątkowe [...] uzasadniają przekonanie, że sprawca grzywny nie uiści.* KK 58,2

(5) *Sąd penitencjarny odwołuje warunkowe zwolnienie, jeżeli zwolniony [...], w okresie próby rażąco narusza porządek prawny...* KKW 160,2

(6) *Organy prowadzące postępowanie karne są obowiązane badać oraz uwzględniać okoliczności przemawiające zarówno na korzyść, jak i na niekorzyść oskarżonego.* KPK 4

(7) *Sędzia penitencjarny wizytuje zakłady karne, areszty śledcze oraz inne miejsca, w których przebywają osoby pozbawione wolności.* KKW 33,1

(8) *Ustawę karną polską stosuje się do sprawcy, który popełnił czyn zabroniony na terytorium Rzeczypospolitej Polskiej...* KK 5

### 3.2. Kolokacje jednostronne

W przeciwieństwie do grupy poprzedniej kolokacje jednostronne mają jednoznacznie określony ośrodek kolokacji oraz kolokat. W ich wypadku element B stanowi dla ośrodka A potwierdzony kolokat (należy do wyrażen przekraczających 10% jego kolokatów), podczas gdy A nie jest potwierdzonym kolokatem B. Tym samym słabsze powiązania obu elementów i związane z nimi słabsze ograniczenia formalno-znaczeniowe sprawiają, że uznanie takich struktur za jednostki wielosegmentowe – względnie potencjalne terminy języka prawnego – jest mniej prawdopodobne. Ośrodkiem kolokacji w takich strukturach bywa (A) czasownik, por. *zwolnić warunkowo, umarzać postępowanie, rozpoznać sprawę, wyznaczać obrońcę*, lub (B) rzeczownik gerundialny, por. *ograniczanie dostępu, postępowanie przygotowawcze, postanowienie o*. Zdecydowanie rzadziej w takiej funkcji wystąpi (C) rzeczownik niegerundialny, por. *orzec nawiązkę, podlega grzywnie, zachowanie tajemnicy, środek przymusu*, czy (D) przymiotnik, przysłówek lub przyimek, por. *działania zbrojne, ponownie popełni*, wyrok *nakazowy, istotne okoliczności, przestępstwo przeciwko*, por.

(9) *Nawiązkę orzeka się na rzecz pokrzywdzonego w wypadkach przewidzianych w przepisach szczególnych.* KW 32

(10) *Sąd umarza postępowanie wykonawcze<sup>11</sup> w razie przedawnienia wykonania kary...* KKW 15,1

<sup>11</sup> W analizowanym przykładzie mamy do czynienia z zażębianiem się kolokacji – [umarza {postępowanie} wykonawcze}.

(11) *Jeżeli wobec skazanego wykonywana jest kolejno kara pozbawienia wolności, a następnie kara orzeczona za wykroczenie albo kara porządkowa lub **środek przymusu** skutkujący pozbawieniem wolności, to w czasie wykonywania tej kary lub środka skazany pozostaje w takim systemie wykonywania kary.* <sup>KKW 99,3</sup>

(12) *Sąd może orzec zakaz prowadzenia pojazdów określonego rodzaju w razie skazania osoby uczestniczącej w ruchu za **przestępstwo przeciwko bezpieczeństwu w komunikacji**.* <sup>KK 42,1</sup>

Jak wspomniano już wyżej (zob. p. 2), większe zróżnicowanie typów dostrzeżga się wśród kolokatów. W tej funkcji mogą bowiem wystąpić: (a) rzeczowniki niegerundialne, por. *użycie **broni**, wykonanie **kary**, zatrzymać **osobę**, **obowiązek** zapłaty, **kara** grzywny*, (b) przymiotki, por. *oskarżyć **o**, zażalenie **na**, postępowanie **w**, wydział **w***, czy (c) czasowniki, por. ***składać** przyrzeczenie, **można** orzec, **należy** doręczyć*. Rzadziej kolokatem bywa (d) rzeczownik gerundialny, por. *wydanie **orzeczenia**, wydaje **postanowienie**, zarządza **wykonanie***, czy (e) przymiotnik lub przysłówek, por. *postępowanie **przygotowawcze**, zarządzi **inaczej**, **niezwłocznie** zawiadamia*, por.

(13) *Sąd, który orzekł karę aresztu **zarządza wykonanie** kary, jeżeli ukarany [...] popełnił podobne [...] wykroczenie.* <sup>KW 44,1</sup>

(14) *...kierownik zakładu **niezwłocznie zawiadamia** o tym właściwą jednostkę Policji...* <sup>KKW 200, 2c</sup>

Warto podkreślenia jest fakt, że w zdecydowanej większości przypadków nadrzędnikiem dystrybucyjnym struktury kolokacyjnej jest ośrodek kolokacji. Zdarzają się jednak wypadki, w których nadrzędnikiem dystrybucyjnym takiej struktury jest kolokat, co jednoznacznie potwierdza tezę, że nie ma bezpośredniej zależności pomiędzy budową grupy składniowej a powiązaniem leksykalnymi wewnątrz kolokacji. Warto również podkreślić, że w grupie drugiej, w której nadrzędnikiem jest kolokat, funkcję tę pełni zwykle czasownik modalny – MÓC, NALEŻY, MOŻNA.

#### 4. Podsumowanie

Prowadzone obserwacje potwierdzają fakt, że język prawny bezpośrednio odwołuje się do modeli składniowo-leksykalnych języka ogólnego. Korzysta jednakże jedynie z niektórych, czego konsekwencją jest ograniczenie zestawu środków komunikacyjnych. Tym samym wskazane jest zidentyfikowanie dla tekstów prawnych dopuszczalnych (zwłaszcza preferowanych) modeli łączliwości składniowo-semantycznej, ujętych np. w postaci słownika kolokacji.

Warto również podkreślić, że zaprezentowane wyniki mają charakter orientacyjny, są bowiem efektem badań pilotażowych, służących głównie wypracowaniu procedur badawczych i metod opisu połączeń występujących w tekstach prawnych. Koniecznym następstwem takich obserwacji jest ich weryfikacja

przeprowadzona – po pierwsze – na szerszym materiale badawczym oraz – po drugie – wykorzystująca inne zautomatyzowane środki wyszukiwania kolokacji, m.in. program MeWeX opracowany w ramach projektu CLARIN (Common Language Resources and Technology Infrastructure<sup>12</sup>). Konieczne jest bowiem uwzględnienie również innych statystycznych algorytmów wyróżniania kolokacji – co było pomijane na dotychczasowym etapie badań – oraz wzięcie pod uwagę niepośrednich połączeń wyrażań, co przy zastosowaniu dotychczasowych narzędzi byłoby bardzo trudne.

## Rozwiązanie skrótów wykorzystywanych kodeksów

KK – *Kodeks karny*, Ustawa z dnia 6 czerwca 1997 r. (z późniejszymi zmianami)

KKW – *Kodeks karny wykonawczy*, Ustawa z dnia 6 czerwca 1997 r. (z późniejszymi zmianami)

KPK – *Kodeks postępowania karnego*, Ustawa z dnia 6 czerwca 1997 r. (z późniejszymi zmianami)

KPW – *Kodeks postępowania w sprawach o wykroczenia*, Ustawa z dnia 24 sierpnia 2001 r. (z późniejszymi zmianami)

KW – *Kodeks wykroczeń*, Ustawa z dnia 20 maja 1971 r. (z późniejszymi zmianami)

## BIBLIOGRAFIA

- Апресян Ю.Д., 2006, *Типы соответствия семантических и синтаксических актантов*, w: *Проблемы типологии и общей лингвистики*, red. В.С. Храковский, С.Ю. Дмитренко, И.М. Заика, Санкт-Петербург, s. 15–27.
- Choduń A., 2004, *Język prawny a język potoczny*, w: *Język – prawo – społeczeństwo*, red. E. Malinowska, Opole, s. 77–86.
- Choduń A., 2006, *Norma językowa a dyrektywy redagowania tekstów aktów prawnych*, w: *Konwencjonalne i formalne aspekty prawa*, red. S. Czepita, Szczecin, s. 47–53.
- Choduń A., 2009, *Uwagi o komunikatywności tekstów aktów prawnych*, w: *Prawo i język*, red. A. Mróz, A. Niewiadomski, M. Pawelec, Warszawa, s. 7–13.
- Gębka-Wolak M., 2014, *Wstępne uwagi o normie składniowej aktów prawnych*, „Poznańskie Spotkania Językoznawcze”, t. 28, s. 25–37, DOI: 10.14746/psj.2014.28.2.
- Gizbert-Studnicki T., 2009, *Postulat jasności i zrozumiałości tekstów prawnych a dostęp do prawa*, w: *Prawo i język*, red. A. Mróz, A. Niewiadomski, M. Pawelec, Warszawa, s. 9–18.

<sup>12</sup> Informacje na temat polskich badań prowadzonych w ramach tego międzynarodowego konsorcjum naukowego oraz dostęp do dotychczas opracowanych narzędzi badawczych znajdują się na stronie internetowej <http://clarin-pl.eu/pl/strona-glowna> (dostęp 15.04.2016).

- Grzelak J., 2010, *Polski język prawa – w perspektywie glottodydaktycznej*, <https://repositorium.amu.edu.pl/jspui/bitstream/10593/352/1/Grzelak.pdf> [dostęp: 25.09.2015].
- Jadacka H., 2010, *Dlaczego nie wszyscy mogą rozumieć teksty prawne*, w: *Prawo, język, etyka*, red. A. Mróz, A. Niewiadomski, M. Pawelec, Warszawa, s. 27–30.
- Karolak S., 1984, *Składnia wyrażen predykatywnych*, w: *Gramatyka współczesnego języka polskiego. Składnia*, red. Z. Topolińska, Warszawa, s. 11–211.
- Kącka-Rodak M., Wichrowska W., 2011, *Poprawność językowa aktów normatywnych – rola redaktora językowego w procesie legislacyjnym*, w: *Prawo, język, media*, red. A. Mróz, A. Niewiadomski, M. Pawelec, Warszawa, s. 131–138.
- L’Homme M.-L., 2005, *Sur la notion de ‘terme’*, „Meta : journal des traducteurs”, t. 50, nr 4, s. 1112–1132, DOI: 10.7202/012064ar.
- Łapa R., 2008, *Znaczenia gramatyczne znominalizowanych grup imiennych we współczesnych tekstach prawnych*, „Język Polski”, LXXXVIII, z. 4–5, s. 280–288.
- Malinowski A., 2006, *Polski język prawny. Wybrane zagadnienia*, Warszawa.
- Moroz A., 2014a, *Granice błędu – norma językowa a teksty prawne*, „Poznańskie Spotkania Językoznawcze”, t. 28, s. 93–107, DOI: 10.14746/psj.2014.28.2.
- Moroz A., 2014b, *Lingwistyczne sposoby kształtowania kulturowego dystansu władzy na przykładzie tekstów prawnych*, w: *Kulturowe konteksty języka*, red. E. Pajewska, Szczecin, s. 211–221.
- Petzel J., 2011, *Status lingwistyczny języka prawnego*, w: *Prawo, język, media*, red. A. Mróz, A. Niewiadomski, M. Pawelec, Warszawa, s. 153–165.
- Preshov K., 2013, *Les collocations dans le discours juridique russe*, „Travaux de Slavistique”, t. 25, s. 175–186.
- Przepiórkowski A., Bańko M., Górski R.L., Lewandowska-Tomaszczyk B. (red.), 2012, *Narodowy Korpus Języka Polskiego*, Warszawa.
- Saloni Z., Świdziński M., 2007, *Składnia współczesnego języka polskiego*, Warszawa.
- Seretan V., 2011, *Syntax-Based Collocation Extraction*, London.
- Zaron Z., 2009, *Problemy składni funkcjonalnej*, Warszawa.
- Zieliński M., 1999, *Języki prawne i prawnicze*, w: *Polszczyzna 2000. Orędzie o stanie języka na przełomie tysiącleci*, red. W. Pisarek, Kraków, s. 50–74.
- Zieliński M., 2004, *Język prawny, język administracyjny, język urzędowy*, w: *Język – Prawo – Społeczeństwo*, red. E. Malinowska, Opole, s. 9–18.
- Zieliński M., 2011, *Mitów o myśleniu o wykładni prawa ciąg dalszy*, w: *Prawo, język, media*, red. A. Mróz, A. Niewiadomski, M. Pawelec, Warszawa, s. 117–126.