

MAŁGORZATA GĘBKA-WOLAK

Uniwersytet Mikołaja Kopernika, Toruń

ORCID: 0000-0003-4587-496X

Korpusy i narzędzia cyfrowe w lingwistycznych analizach problemów prawnych. Jednostki tekstu prawnego w ustawie „Prawo własności przemysłowej”

1. Wprowadzenie

Przedmiotem uwagi w artykule jest warstwa leksykalna tekstu prawnego¹, analizowana w ramach koncepcji metodologicznej, której centralnym elementem myślowym jest pojęcie jednostki tekstu prawnego (JTP)².

1.1. Koncepcja JTP – podstawy teoretyczne oraz zastosowanie

Jednostką tekstu prawnego jest nazywane wyrażenie językowe dysponujące równocześnie dwiema cechami: (1) jest jednostką języka³, tj. obiektem będącym skojarzeniem

¹ Tekst prawny to tekst konkretnego aktu prawodawczego, tj. ustawy, rozporządzenia, zarządzenia, uchwały (Zieliński, 1999, s. 50; 2004, s. 11, 13). Tekst prawny jest zarazem wykładnikiem prawnej odmiany języka, tj. w nim uzewnętrznia się językowa specyfika tej specjalistycznej odmiany polszczyzny.

² Zob. Gębka-Wolak, Moroz (2019).

³ Koncepcja JTP została wypracowana w nawiązaniu do teorii jednostki języka i gramatyki operacyjnej Andrzeja Bogusławskiego (1988). Teoria ta jest twórczo rozwijana przez wielu językoznawców, m.in. Macieja Grochowskiego (1982, 1983, 1993, 1997, 2001, 2002, 2012), Magdalenę Danielewiczową (2012, 2014, 2016, 2021) oraz Adama Dobaczewskiego (2002, 2009, 2018). Znajduje ponadto zastosowanie w ogólnych i specjalistycznych opracowaniach leksykograficznych (Bogusławski i Danielewiczowa, 2005; Bogusławski i Wawrzyńczyk, 1993; Dobaczewski, Sobotka i Żurowski, 2018), jednak wcześniej nie była wykorzystywana w badaniach tekstów prawnych, a – jak starano się wykazać w monografii M. Gębki-Wolak i A. Moroz (2019) – może stanowić metodologiczną podstawę do opisu warstwy leksykalnej prawnej odmiany polszczyzny.

signifiant i *signifié* albo na mocy reguł systemowych, albo na mocy apriorycznej decyzji jednostkowej, przebiegającej zgodnie z procesami terminologizacji i dekretowanej za pomocą definicji legalnej lub specjalistycznej; (2) jest stosowane w tekstach prawnych do wyrażania treści specjalistycznych lub ogólnych i jako takie służy działaniom związanym ze stanowieniem prawa (Gębka-Wolak, Moroz, 2019, s. 97–98).

Koncepcja JTP opiera się na założeniu tożsamości i różnicy konwencji językowych warunkujących budowanie tekstów prawnych, będących wykładnikami prawnej odmiany języka, w stosunku do konwencji leżących u podstaw tworzenia komunikatów językowych identyfikowanych jako należące do ogólnej odmiany języka polskiego. Zgodnie z założeniem tożsamości język prawny nie jest odrębnym kodem, niezależnym od reguł języka polskiego, lecz jego odrębną realizacją. Jest specjalistyczną, sztucznie kształtowaną realizacją kodu językowego⁴. Wykorzystuje jego jednostki i mechanizmy, w szczególności system cech diakrytycznych tworzących zestaw fonemów, mechanizm fleksyjno-składniowy służący do kodowania związków syntagmatycznych, kategorie gramatyczne i semantyczne, reguły derywacji, łączliwości i linearyzacji. Zgodnie z założeniem różnicy język prawny stanowi natomiast przykład funkcjonalnej specjalizacji użycia kodu, kształtowania go „zgodnie z wolą twórców” (Grochowski, 2017, s. 12)⁵; od ich decyzji zależy m.in. sposób wiązania w postaci jednostek języka określonego znaczenia z określonym kształtem, przyporządkowywania „signifiant do signifié”. W konsekwencji największym przeobrażeniom podlega podsystem leksykalno-składniowy: powstają nowe jednostki, inne są modyfikowane w planie formalnym i/ lub semantycznym. Zmianie podlega także sieć powiązań leksykalnych, ustanawiane są bowiem inne paradygmatyczne i syntagmatyczne związki leksykalne (Gębka-Wolak, Moroz, 2019, s. 45–49, 105–152).

Docelowo koncepcja JTP stanowić ma metodologiczną podstawę do przygotowania leksykograficznego opisu prawnej odmiany polszczyzny, opracowania mocno osadzonego na gruncie językoznawczym, a jednocześnie w sposób systematyczny uwzględniającego semantyczno-składniową specyfikę prawnej odmiany języka, ze szczególnym uwzględnieniem cech utrudniających rozumienie tekstów prawnych przez niespecjalistów. Próbną wersja „Słownika jednostek tekstu prawnego”, obejmująca założenia i zasady opisu oraz przykładowe hasła, została zawarta w monografii „Jednostka tekstu prawnego w ujęciu teoretycznym i praktycznym” (Gębka-Wolak, Moroz, 2019, s. 153–206). Aby zapewnić spójność terminologii prawnej, baza materiałowa „Słownika” została ograniczona do sześciu kodeksów z zakresu prawa karnego (Gębka-Wolak, Moroz, 2019, s. 51–52, 157–158). Z tego samego powodu kontynuacja opisu prawnej odmiany polszczyzny powinna uwzględniać przyjmowane na gruncie prawa grupowanie norm

⁴ Odwołując się do wprowadzonej przez A. Bogusławskiego (2008, s. 20–26) dychotomii: *narzędzie – działanie na narzędziu*, można powiedzieć, że język prawny jest działaniem na narzędziu, jakim jest język naturalny zdeponowany w świadomości danej społeczności.

⁵ Przez kształtowanie należy rozumieć, że odmiana prawna jest tworzona „w sposób świadomy i celowy przez legislatorów w wyniku wieloletniego i powtarzalnego konstruowania wypowiedzi opisujących stanowione przez nich dyrektywy postępowania” (Malinowski, 2006, s. 235).

w poszczególne działy, gałęzie, zwane także dziedzinami⁶. W praktyce oznacza to, że docelowe opracowanie będzie się składało z wielu modułów, przygotowywanych etapami, począwszy od zestawienia i zanalizowania dziedzinowych podkorpusów tekstów prawnych, przez identyfikację JTP i ustalenie ich właściwości do zebrania wyników analiz w postaci haseł słownikowych.

1.2. Przedmiot i cel badania pilotażowego

Mając na uwadze to, że koncepcja JTP została skonstruowana na podstawie oglądu ograniczonego zbioru tekstów prawnych, zasadniczy etap prac leksykograficznych należy poprzedzić pilotażowymi badaniami tekstów prawnych należących do różnych działów i gałęzi prawa. W ten sposób możliwa będzie weryfikacja trafności hipotez sformułowanych na podstawie analizy wyjściowego korpusu tekstów z zakresu prawa karnego oraz przydatności stosowanych narzędzi, a także uściślenie procedur badawczych. Pierwsze tego rodzaju badanie pilotażowe, którego wyniki są prezentowane w niniejszym artykule, zostało przeprowadzone na materiale ustawy z 30 czerwca 2000 roku „Prawo własności przemysłowej” (tekst jedn. Dz.U. z 2021 r. poz. 213), wchodzącej w skład grupy ustaw z dziedziny prawa własności intelektualnej (Michniewicz, 2022, s. XIII). Wybór tego tekstu prawnego ma dwojaką motywację: 1) reprezentuje gałąź prawa o innym charakterze i przedmiocie regulacji niż prawo karne⁷; 2) jest zarazem przedmiotem obserwacji ze względu na równoległe prowadzone badania dotyczące statusu wyrażen językowych odnoszących się do objętych ochroną prawną słownych znaków towarowych.

Omawiane w artykule badanie pilotażowe miało na celu sprawdzenie przydatności wybranego narzędzia cyfrowego, aplikacji webowej Korpusomat⁸, do analizy specyfiki językowej tekstu prawnego i opracowania „Słownika jednostek tekstu prawnego”, w szczególności zaś wspomaganie identyfikacji poszczególnych klas JTP.

Zbiór obiektów leksykalnych objętych wspólną etykietą JTP jest niejednorodny, dlatego do jego opisu został wykorzystany aparat obejmujący klasy wyodrębnione na podstawie kryteriów⁹: (a) aprioryczności procesu definiowania; (b) istnienia definicji legalnej; (c) odmienności formalnej i/lub znaczeniowej jednostek tekstowych w stosunku do ogólnych jednostek systemowych, opartej na modyfikacjach zindywidualizowanych

⁶ Na temat porządkowania prawa na podstawie kryterium przedmiotowego zob. np. Muras (2020, s. 41, 52–57).

⁷ Inaczej niż prawo karne, które należy do podstawowych („samowystarczalnych”) gałęzi prawa, prawo własności intelektualnej jest traktowane jako kompleksowa gałąź prawa, tj. gałąź, która korzysta z norm należących do kilku innych gałęzi prawa, m.in. prawa cywilnego, administracyjnego, gospodarczego, pracy (Michniewicz, 2022, s. 4–5).

⁸ Wybór Korpusomatu wynika z tego, że aplikacja ta z powodzeniem została wykorzystana do utworzenia i badania korpusu tekstów z dziedziny prawa karnego. W zakresie wydobywania terminologii dziedzinowej Korpusomat korzysta zaś z aplikacji TermoPl (Kieraś, Kobyliński, 2021, s. 54).

⁹ Omawiane tu z konieczności w skrócie kryteria i wynikająca z nich klasyfikacja JTP zostały szczegółowo przedstawione w monografii M. Gębki-Wolak i A. Moroz (2019, s. 98–104).

lub zgodnych z mechanizmami gramatycznymi istniejącymi w polszczyźnie. Na tej podstawie zostało wyróżnionych pięć klas obiektów:

(1) podstawowe warianty jednostek języka (PWJJ) – jednostki, którym nie przypisano znaczenia w sposób aprioryczny, przeniesione na grunt polszczyzny prawnej bez poddawania procesom modyfikacji formalnej lub semantycznej;

(2) systemowe warianty jednostek języka (SWJJ) – jednostki, którym nie przypisano znaczenia w sposób aprioryczny, przeniesione na grunt polszczyzny prawnej z wykorzystaniem systemowych mechanizmów (derywacja) dostosowujących ich stronę formalno-znaczeniową do użycia w tekstach prawnych;

(3) pozasystemowe warianty jednostek języka (PSWJJ) – jednostki, którym nie przypisano znaczenia w sposób aprioryczny, przeniesione na grunt polszczyzny prawnej z wykorzystaniem zindywidualizowanych (jednostkowych) mechanizmów dostosowujących ich stronę formalno-zaczeniową do użycia w tekstach prawnych;

(4) legalne jednostki terminologiczne (LJT) – jednostki, którym przypisano znaczenie w sposób aprioryczny, dysponujące na gruncie prawnym definicjami legalnymi;

(5) specjalne jednostki terminologiczne (SJT) – jednostki, którym przypisano znaczenie w sposób aprioryczny, niedysponujące na gruncie prawnym definicjami legalnymi.

Ze względu na aprioryczność procesu definiowania klasy (4) i (5) można zarazem uznać za terminologię dziedzinową, której identyfikacja w ramach tworzenia „Słownika jednostek tekstu prawnego” w szczególności powinna być wspomagana narzędziami cyfrowymi.

Materiał do badań – tekst ustawy „Prawo własności przemysłowej” – został zgromadzony w formie elektronicznego pliku zapisanego w formacie pdf, a także w postaci binarnego korpusu (dalej: korpus Pwp) wygenerowanego za pomocą Korpusomatu (zob. Kieraś, Kobyliński, 2021), czyli „aplikacji webowej służącej do samodzielnego gromadzenia i indeksowania znakowanych korpusów językowych z tekstów dostarczonych przez użytkownika” (Kieraś, Kobyliński, 2021, s. 49). W badaniach została wykorzystana wersja aplikacji zawierająca poza standardowymi fleksyjnymi warstwami automatycznego znakowania tekstu także parsowanie zależnościowe i rozpoznawanie jednostek nazewniczych. Korpusomat w tej wersji jest wyposażony ponadto w narzędzia do przedstawiania podstawowych statystyk, w tym wykazu słownictwa charakterystycznego dla badanego korpusu, generowanego przez aplikację TermoPI¹⁰.

2. Badanie korpusowe

Korpus Pwp wygenerowany za pomocą Korpusomatu zawiera 49 550 segmentów, co oznacza, że z perspektywy automatycznego przetwarzania danych jest to korpus mały. W odniesieniu do korpusu specjalistycznego wielkość nie jest jednak parametrem

¹⁰ Aplikacja jest dostępna pod adresem: <https://korpusomat.pl>. W badaniu została wykorzystana wersja z uzupełnieniami datowanymi na 21.01.2021, zob. <https://korpusomat.pl/blog/posts/sentiment.html> (dostęp: 1.04.2023).

o kluczowym znaczeniu dla interpretacji danych, zwłaszcza że badanie ma charakter pilotażowy. Ręczne przeszukiwanie blisko pięćdziesięciotysięcznego zbioru wyrażań byłoby czynnością czasochłonną, a zarazem obciążoną ryzykiem pominięcia w analizie takich właściwości badanego materiału, które ujawniają się dopiero podczas porównania zbiorczych danych uzyskanych na podstawie różnych parametrów gramatycznych, co zapewniają narzędzia, w które wyposażony jest np. Korpusomat. Jest natomiast oczywiste, że powodzenie realizacji celu badań wspomaganych tego rodzaju narzędziami zależy od zaplanowania odpowiedniej procedury przeszukiwania korpusu.

Jeśli chodzi o badanie opisywane w artykule, podstawowym zadaniem wykorzystania narzędzi korpusowych było wspieranie identyfikacji – po pierwsze – poszczególnych klas JTP, zwłaszcza LJT i SJT, a więc specjalistycznej terminologii prawnej, po drugie – formalnych modyfikacji jednostek polszczyzny ogólnej, którym mogą podlegać SWJJ, PSWJJ, LJT i SJT. Badanie korpusu zaplanowano w dwóch etapach: pierwszy dostarczał ogólnych informacji o właściwościach gramatycznych badanego materiału i podstawowych danych frekwencyjnych, drugi bezpośrednio służył identyfikacji poszczególnych typów JTP oraz formalnych przekształceń jednostek leksykalnych należących do polszczyzny ogólnej (LJW). Wykorzystane zostały narzędzia do przeszukiwania materiału i statystyczne, dostępne w zakładkach „Odpytaj korpus” i „Statystyki korpusu”.

2.1. Ogólna charakterystyka materiału

Moduł „Statystyki korpusu” dostarcza skróconej (w oknie „Lista frekwencyjna”) i pełnej (po pobraniu) listy frekwencyjnej ze względu na lematy słów¹¹. Zawiera ona również informacje na temat przynależności słów do klas gramatycznych, co pozwala się zorientować, jakie części mowy i w jakiej proporcji występują w badanym materiale. Przeprowadzona automatycznie identyfikacja klas gramatycznych jest na ogół poprawna. Błędna kwalifikacja dotyczy głównie słów homonimicznych. Na przykład ciągi *czas*, *sposób* i *to* w kilku kontekstach zostały błędnie uznane za predykatywy (por. *czas trwania patentu*, *sposób badania tych właściwości*, *sprawozdanie to zostanie złożone do akt*). Z kolei ciąg *mowa* we wszystkich wyrażeniach typu *o których mowa w art.* zamiast do klasy rzeczowników został zaliczony do wykrzykników. Za wykrzyknik został także uznany przyimek *o* przy kilku wystąpieniach rzeczowników *wnioszek* (por. *złożonych wnioskach o ustalenie*) i *sprawa* (por. *rozstrzyga sprawę o: 1) unieważnienie patentu*). Przyimek *za* został zaś w kilku kontekstach zaliczony do klasy partykuł (por. *Urząd Patentowy uzna sprzeciw za wycofany*). W klasie partykuł liczba błędnych interpretacji jest zresztą stosunkowo duża. Nieprawidłowa kwalifikacja dotyczy np. kilku wystąpień wyrażenia *chyba*, które zostało opisane jako partykuła, a jest segmentem spójnika *chyba, że* (por. [...] *jako nazwy handlowej lub nazwy przedsiębiorstwa albo*

¹¹ Z opisu aplikacji wynika, że pobrana lista zawiera pierwszych 1000 najczęściej występujących lematów, zob. <https://korpusomat.readthedocs.io/pl/latest/statystyki.html> (04.2023). Ze względu na niewielką objętość badanego korpusu pobrana lista frekwencyjna zawiera wszystkie segmenty.

jako części tych nazw, chyba że pozostaje ono bez wpływu na możliwość odróżnienia towarów w obrocie [...]), wyrażenia *może*, które jest formą czasownikową, a nie partykułą (por. *Pracownik, który złożył egzamin wymagany zgodnie z ust. 3, może uzyskać uprawnienia eksperta z pominięciem okresu asesury.*), wyrażenia *i*, które jest spójnikiem, a nie partykułą (por. *Roszczenia dotyczące wynalazków, wzorów użytkowych, wzorów przemysłowych i topografii układów scalonych*).

W badanym tekście najliczniejszy jest zbiór słów rzeczownikowych: rzeczowniki (906) i gerundia (277)¹² stanowią 36,5% wszystkich lematów wydzielonych w korpusie, co jest zgodne z rozkładem części mowy w polszczyźnie ogólnej i prawnej. Wynik ten potwierdza odnotowaną przez Andrzeja Malinowskiego (2006, s. 66) ogólną właściwość tekstów prawnych, polegającą na znacznie większym udziale rzeczowników niż w tekstach popularnonaukowych, drobnych wiadomościach prasowych i publicystyce, w których wynosi on około 30%. Wśród 100 najczęstszych słów (lematów) badanego korpusu 37 to rzeczowniki i gerundia. Na czele listy są wyrażenia: *prawo* (617 wystąpień), *znak* (444 wystąpień), *urząd* (426 wystąpień), *wynalazek* (278 wystąpień), *wzór* (274 wystąpień). Dla porównania rzeczownikami o najwyższej frekwencji w korpusie składającym się z sześciu kodeksów reprezentujących prawo karne są wyrażenia: *kara*, *sąd*, *postępowanie*, *osoba*, *przestępstwo*, *przepis*.

Warte uwagi są ponadto wyniki analizy frekwencji poszczególnych form wyrazowych czasowników, zob. Tabelę 1. Jest to możliwe dzięki temu, że w Korpusomacie części mowy są oznaczane w oparciu o zestaw wartości uwzględniający zróżnicowane typy budowy form czasownikowych.

Podane w Tabeli 1. dane liczbowe należy traktować jako orientacyjne, gdyż automatyczna identyfikacja typów form czasownikowych nie zawsze jest poprawna. Na przykład rzeczownik *dział*, będący nazwą części aktu prawnego (por. *DZIAŁ II Przepisy przejściowe i końcowe*), został zinterpretowany jako pseudoimiesłów. Spośród 20 przykładów zidentyfikowanych jako zawierające rozkaźniki żaden takowej formy nie zawierał. Były to bowiem spójniki *lub*, *bądź* (por. *wzór jest zawarty bądź zastosowany; prawa ochronnego lub odpowiednio prawa z rejestracji*), segment *spraw* w przyimku *do spraw*, a także forma rzeczownika *prawo* (por. *przewidujących ochronę praw do odmiany roślin*). Z kolei wśród 4 lematów zidentyfikowanych jako predykatywy, tj. *można*, *to*, *sposób*, *czas*, jest nim tylko *można*. Pomyłki zdarzają się także w lematyzacji formy przyszłej *być*. W programie wystąpienia te są traktowane jako realizacje leksemu *być*, co nie jest zgodne ze stanem faktycznym. Wyszukane zostały bowiem przykłady z formami czasu przyszłego czasowników *przysługiwać*, *wykorzystywać*, *prowadzić*, *używać* oraz konstrukcje bierne: *będzie dochodzone*, *będą brane pod uwagę*, *będzie ograniczony*.

¹² Wartości klas gramatycznych uwzględniane w Korpusomacie są inne niż w klasyfikacji tradycyjnej, bardziej szczegółowe. Mamy możliwość wyszukania 34 klas, w tym poza standardowymi także takie jak: rzeczownik deprecjatywny, przymiotnik przyprzymiotnikowy, przymiotnik poprzyimkowy, przymiotnik predykatywny, skrót. Jest to zarazem standard opisu przyjęty w innych narzędziach cyfrowych przygotowanych do przetwarzania polszczyzny, np. w NKJP (zob. Kieraś i Kobyliński, 2021, s. 50; Przepiórkowski i in., 2012, s. 62–63).

Typy form czasownikowych	Liczba wystąpien w korpusie Pwp	Liczba form podstawowych (lematów)	Przykłady lematów, których formy mają najwyższą frekwencję
forma nieprzeszła	1864	co najmniej 100 ¹³	móc, być, stosować, podlegać, wydawać, uważać, mieć
bezokolicznik	503	co najmniej 100	być, żądać, wystąpić, odpowiadać, zawierać, zostać, dołączyć
pseudoimiesłów	327	67	zostać, być, móc, nastąpić, dziać, uzyskać
winien	59	1	powinien
imiesłów przysłówkowy współczesny	47	27	uwzględniać, być, orzekać, wyznaczać, działać, liczyć
predykatyw	43	4	można, to, sposób
bezosobnik	30	14	udzielić, stwierdzić, wnieść, dokonać, korzystać
rozkaznik	20	5	lubić, być
forma przyszła <i>być</i>	7	1	być
imiesłów przysłówkowy uprzedni	1	1	począć
aglutynat <i>być</i>	0	0	

Tabela 1. Obciążenie funkcjonalne form czasownikowych

Pomimo wyróżnionych niedoskonałości automatycznej identyfikacji klas gramatycznych dane wyszukane w Korpusomacie wyraźnie wskazują na różnicę obciążenia funkcjonalnego form czasownikowych w badanym tekście prawnym. Formy nieprzeszłe stanowią 64,7% wszystkich form czasownikowych użytych w korpusie Pwp. Przewagę form nieprzeszłych należy wiązać z zasadami techniki prawodawczej, zgodnie z którymi wypowiedzi dyrektywne są formułowane w trybie oznajmującym, najczęściej w czasie teraźniejszym lub przyszłym prostym. Czasowniki w czasie przeszłym nie pełnią zaś funkcji dyrektywnej (Malinowski, 2012, s. 109–110). W badanym tekście pseudoimiesłowów jest blisko sześć razy mniej (stanowią 11,4% użytych form czasownikowych) niż form nieprzeszłych. Dodajmy, że ciągi identyfikowane w Korpusomacie jako pseudoimiesłowy obejmują wyrażenia o zróżnicowanej budowie. Są to: formy czasu przeszłego w stronie czynnej (np. *sprawca działań w dobrej wierze*), formy czasu przeszłego stanowiące składnik konstrukcji (strony) biernej (np. *tłumaczenia zostały złożone*), formy trybu przypuszczającego stanowiące składnik konstrukcji biernej (np. *dochodzenia byłyby należne*), formy trybu przypuszczającego w stronie czynnej (np. *korzystanie wkraczałoby w zakres patentu*).

¹³ Korpusomat do 100 ogranicza liczbę wyświetlanych lematów.

Uzupełnieniem informacji o morfologicznych właściwościach badanego tekstu prawnego jest ustalenie, jakie części mowy w nim nie są reprezentowane. Dane uzyskujemy za pomocą serii zapytań o części mowy, dostępnych w module „Odpytaj korpus”, np. [pos="(ppron12)"] – pytanie o zaimki nietrzeciosobowe, [pos="(depr)"] – pytanie o rzeczowniki deprecjatywne, [pos="(brev)"] – pytanie o skróty. Z wyszukiwań przeprowadzonych w korpusie Pwp wynika, że nie ma w nim rzeczowników deprecjatywnych, przymiotników poprzyimkowych i predykatywnych, a zaimki nietrzeciosobowe są reprezentowane przez jeden przykład – cytat ze ślubowania rzeczownika patentowego.

Korpusomat ma wbudowane narzędzia do wyszukiwania słów kluczowych. Lista takich słów, dostępna w module „Statystyki korpusu”, jest tworzona nie tylko z uwzględnieniem parametrów frekwencyjnych, ale także „na podstawie porównania listy frekwencyjnej korpusu z listą frekwencyjną korpusu referencyjnego. Korpusem referencyjnym w tym przypadku jest milionowy korpus NKJP¹⁴. W korpusie Pwp najczęstsze słowa kluczowe w znacznym stopniu pokrywają się ze słowami o najwyższej frekwencji, co pokazuje zacienienie kolumn w Tabeli 2. W pierwszej i drugiej kolumnie Tabeli 2. zostało podanych po 15 słów o najwyższej frekwencji lub kluczowych (niebędących zaimkami i przyimkami). W kolumnie trzeciej – dla porównania – zostało zamieszczonych 15 najczęstszych rzeczowników.

Lista frekwencyjna z Pwp	Lista słów kluczowych z Pwp	Lista frekwencyjna rzeczowników z Pwp
<i>artykuł</i>	<i>prawo</i>	<i>prawo</i>
<i>prawo</i>	<i>patentowy</i>	<i>znak</i>
<i>ustęp</i>	<i>znak</i>	<i>urząd</i>
<i>patentowy</i>	<i>urząd</i>	<i>wynalazek</i>
<i>znak</i>	<i>towarowy</i>	<i>wzór</i>
<i>urząd</i>	<i>ochronny</i>	<i>zgłoszenie</i>
<i>towarowy</i>	<i>wynalazek</i>	<i>patent</i>
<i>ochronny</i>	<i>wzór</i>	<i>przepis</i>
<i>mowa</i>	<i>zgłoszenie</i>	<i>osoba</i>
<i>móc</i>	<i>patent</i>	<i>sprawa</i>
<i>wynalazek</i>	<i>przepis</i>	<i>rejestracja</i>
<i>wzór</i>	<i>udzielić</i>	<i>wniosek</i>
<i>zgłoszenie</i>	<i>rejestracja</i>	<i>postępowanie</i>
<i>patent</i>	<i>postępowanie</i>	<i>termin</i>
<i>przepis</i>	<i>stosować</i>	<i>decyzja</i>

Tabela 2. Słowa o najwyższej frekwencji w Pwp

¹⁴ Zob. dostępny on-line opis aplikacji Korpusomat: <https://korpusomat.readthedocs.io/pl/latest/statystyki.html> (dostęp: 1.04.2023).

Zwróćmy uwagę na to, że słowa kluczowe oraz te o wysokiej frekwencji, szczególnie rzeczowniki, wyszukane w dziedzinowych korpusach tekstów prawnych, tj. uwzględniających przyjmowane na gruncie prawa grupowanie norm w poszczególne działy, gałęzie czy dziedziny, są zarazem leksykalnymi wyróżnikami (identyfikatorami) tejsze dziedziny/gałęzi prawa. Takiego charakteru nie mogą mieć natomiast słowa o najwyższej frekwencji w korpusie tekstów prawnych, który był podstawą badań A. Malinowskiego (2006, s. 56–60), gdyż korpus ten obejmował próbki wielu aktów normatywnych opublikowanych do 1 stycznia 2005 roku. Na liście frekwencyjnej zamieszczonej w opracowaniu Malinowskiego najwyższą frekwencję poza znajdującymi się w pierwszej dziesiątce przyimkami i spójnikami mają słowa: *być, mowa, dzień, móc, praca, sprawa, przepis, osoba, właściwy, minister* (zob. Malinowski, 2006, s. 70).

W funkcji leksykalnego wyróżnika dziedziny/gałęzi prawa jeszcze lepiej sprawdzają się wyrażenia nazywane w Korpusomacie „Słownictwem charakterystycznym”, wyekscerpowane dzięki zastosowaniu specjalnych algorytmów do automatycznego wydobywania z korpusów terminologii dziedzinowej, dostępnych w aplikacji TermoPl. Jest to efekt zastosowania na potrzeby narzędzia komputerowego bardziej sformalizowanej definicji terminologii dziedzinowej. Zgodnie z założeniami przyjętymi przez zespół opracowujący aplikację TermoPl terminy dziedzinowe to n-gramowe frazy rzeczownikowe o stosunkowo wysokiej frekwencji w tekstach dziedzinowych (zob. Marciniak, Mykowiecka i Rychlik, 2017, s. 64–66). Spośród 25 wyrażeń, które w Korpusomacie znalazły się na liście słownictwa charakterystycznego¹⁵, 18 wyrażeń można uznać za leksykalne wyróżniki dziedzinowe. Są to bowiem nazwy obiektów będących przedmiotem regulacji prawa własności przemysłowej lub działań związanych z regulacją prawną¹⁶. Wyrażenia te podane zostały w pierwszej kolumnie Tabeli 3. W kolumnie drugiej są wyrażenia tematycznie nierelevantne dla badanego tekstu prawnego.

Wyrażenia z listy „Słownictwo charakterystyczne”	
relevantne dla prawa własności przemysłowej	nierelevantne dla prawa własności przemysłowej
<i>urząd patentowy</i>	<i>kancelaria sejmu</i>
<i>znak towarowy</i>	<i>Rzeczpospolita Polska</i>
<i>prawo ochronne</i>	<i>ponowne rozpatrzenie sprawy</i>
<i>wzór przemysłowy</i>	<i>osoba trzecia</i>
<i>wzór użytkowy</i>	<i>wydanie decyzji</i>

¹⁵ Ograniczenie dotyczy także listy pobieranej po naciśnięciu przycisku „Pobierz”. Więcej wyników można uzyskać w odrębnym badaniu w aplikacji TermoPL. Na liście wyników otrzymanych w wyniku analizy w aplikacji webowej znalazło się 901 wyrażeń co najmniej 2-gramowych.

¹⁶ Wyrażenia te są np. objaśniane lub komentowane w rozmaitych kompendiach prawnych (zob. np. Barczewska i Kowalska, 2019; Michniewicz, 2022), a także w słowniczku terminów publikowanym na stronie Urzędu Patentowego RP, zob. <https://uprp.gov.pl/pl/slownik-terminow> (dostęp: 1.04.2023).

Wyrażenia z listy „Słownictwo charakterystyczne”	
relewantne dla prawa własności przemysłowej	nierelwantne dla prawa własności przemysłowej
<i>dodatkowe prawo ochronne</i>	<i>prezes Rady Ministrów</i>
<i>prezes Urzędu Patentowego</i>	<i>wyznaczony termin</i>
<i>oznaczenie geograficzne</i>	
<i>udzielone prawo</i>	
<i>udzielenie patentu</i>	
<i>uzyskanie patentu</i>	
<i>rejestracja wzorów przemysłowych</i>	
<i>zgłoszenie wynalazku</i>	
<i>układ scalony</i>	
<i>topografia układu scalonego</i>	
<i>udzielone prawo ochronne</i>	
<i>wcześniejszy znak towarowy</i>	
<i>zgłoszenie wynalazku</i>	

Tabela 3. Słownictwo charakterystyczne dla korpusu Pwp – lista z Korpusomatu

Wyrażenia odnotowane na liście słownictwa charakterystycznego są zarazem podstawowymi choć niejedynymi kandydatami na terminy dziedziczne, czyli JTP z klas LJT i SJT.

2.2. Identyfikacja poszczególnych klas JTP

2.2.1. Procedura falsyfikacyjna

Zastosowanie w module „Odpytaj korpus” wyszukiwania z atrybutem „ign” daje możliwość sprawdzenia, czy badany korpus zawiera formy wyrazowe nieobecne w „Słowniku gramatycznym języka polskiego”, co daje podstawę do wysuwania hipotez o neologicznym statusie wyrażenia, dla których wartość tego parametru równa się „true” (zob. Kieraś, Kobyliński, 2021, s. 51). Mimo że zbiór wyników dla zapytania [ign="(true)"] nie jest pusty, o korpusie Pwp można powiedzieć, że jego warstwa leksykalna bazuje na leksyce odnotowanej w „Słowniku gramatycznym”, a więc leksyce o pewnym stopniu ustabilizowania w polszczyźnie współczesnej. Te bowiem ciągi, które zostały wskazane jako formy nierozpoznane (56 przykładów), nie były formami wyrazowymi, lecz ich przypadkowymi fragmentami, np. fragmentami słów przenoszonych do kolejnego wiersza (np. *wyna-*, *wcześ-*, *-niejszego*), skrótami bez kropki lewostronnie dosuniętej do znaku (*sygn*, *późn*). Nie zostały rozpoznane albo w wyniku błędnej edycji tekstu w pliku wyjściowym, albo w wyniku pomyłki na etapie tworzenia korpusu binarnego.

Należy jednak podkreślić, że ze względu na stosowane zabiegi kształtujące prawną odmianę polszczyzny, które stały się zresztą podstawą do wypracowania aparatu JTP zawierającego pięć klas jednostek, otrzymany wynik nie jest równoznaczny z uznaniem, że leksyka Pwp jest w pełni zbieżna z zasobem leksykalnym polszczyzny odnotowanym w „Słowniku gramatycznym”. Ustalenie, które wyrażenia są PWJJ, a które mają inny status, wymaga dokładniejszych analiz, w tym także kolejnych wyszukiwań w korpusie Pwp.

Otrzymany wynik badania korpusowego podsunął natomiast pomysł, by do identyfikacji poszczególnych klas JTP wykorzystać następującą procedurę o charakterze falsyfikacyjnym. Zakładamy, że w wyniku dokonanej w Korpusomacie automatycznej segmentacji i lematyzacji wyrażen oraz wyłączeniu z otrzymanego zbioru wyrażen niebędących leksemami (znaki interpunkcyjne, fragmenty form wyrazowych) oraz cyfr i liter będących numerami kolejnych elementów struktury aktu prawnego (numery działów, artykułów, paragrafów, ustępów, punktów) otrzymaliśmy zbiór PWJJ. Jeśli w wyniku przeprowadzonych kolejnych badań korpusowych i analiz językoznawczych nie znajdziemy danych podważających ten status, kwalifikacja wstępna pozostanie w mocy.

Pełna lista frekwencyjna wygenerowana w programie Korpusomat dla korpusu Pwp obejmowała 3242 lematy. Po wyeksportowaniu do programu Excel poddana została sortowaniu wg kryterium klasy gramatycznej, następnie filtrowaniu – usunięto wyrażenia oznaczone jako „romandig”, „dig”, „xxx”, „interp”, „brev”, „frag”, a także nieoznaczone jako zawierające znaki interpunkcyjne i cyfry lub litery ciągi typu <1, <Art. oraz ciągi alfanumeryczne typu 1526a. Po operacji filtrowania lista zmniejszyła się do 2678 lematów, którym wstępnie został przypisany status PWJJ. Następnym etapem zastosowanej procedury była weryfikacja statusu PWJJ wspomagana narzędziami dostępnymi z poziomu aplikacji Korpusomat. Biorąc pod uwagę kryteria wykorzystane do klasyfikacji JTP, w szczególności starano się ustalić, czy na wstępnej liście PWJJ są lematy:

- 1) którym przypisano definicje legalne bądź specjalistyczne lub które są segmentami wyrażen mających takie definicje;
- 2) których kształt lub sposób użycia w kontekście wskazywałyby na to, że mamy do czynienia z inną jednostką niż znana w polszczyźnie ogólnej.

2.2.2. LJT

W przygotowanym korpusie stosunkowo najłatwiej przebiegała identyfikacja jednostek mających definicje legalne równościowe klasyczne – trójczłonowe – lub nieklasyczne – zakresowe, przez wyliczenie (Malinowski, 2006, s. 166–175). W tego rodzaju definicjach są bowiem stosowane wyrażenia, które mogą być podstawą łatwych do sformułowania w Korpusomacie wyrażen wyszukiwawczych. Poniżej zostały przedstawione formuły wyszukiwawcze, których użycie zwróciło niepuste zbiory wyników.

1) [(pos="(subst)"&case="(inst"))][[base="(być)"]

Formuła ta uwzględnia postać leksykalną elementu łączącego definiens z definiendum, którym jest forma czasownika *być*, najczęściej *jest*, a także formę i szyk definiensa. Definiens to dwuskładnikowa¹⁷ narzędnikowa fraza rzeczownikowa, na linii tekstu umieszczona lewostronnie względem elementu łączącego. Formuła pozwala wyszukać LJT zawarte w definicjach klasycznych, zob. (1), lub zakresowych, zob. (2):

(1) Wynalazkiem tajnym jest wynalazek stanowiący informację niejawną, której nadano klauzulę: „ściśle tajne”, „tajne”, „poufne” lub „zastrzeżone”;

(2) Karami dyscyplinarnymi są: 1) upomnienie; 2) nagana [...].

Zidentyfikowane LJT to *wynalazek tajny* w (1) i *kara dyscyplinarna* w (2).

2) [base="(rozumieć)"|base="(rozumienie)"]

Druga formuła wyszukiwawcza pozwala wyszukać LJT definiowane przy pomocy konstrukcji zawierających formy czasownika *rozumieć* lub jego rzeczownikowego derywatu *rozumienie*: *rozumie się przez to, należy rozumieć, przez ... należy rozumieć ..., w rozumieniu (czego, np. przepisów, ustawy, ust.), pod pojęciem ... rozumie się*. Zastosowanie formuły w korpusie Pwp zwróciło najwięcej przykładów, bo 56. Co więcej, większość zawierała jedno- lub wielowyrazowe LJT, np. *projekt racjonalizatorski, osoba zagraniczna, przedsiębiorca, projekt wynalazczy, stan techniki, konwencja paryska, projekt racjonalizatorski, postać elektroniczna zgłoszenia, towar, zwykłe używanie, topografia układu scalonego, znak towarowy, międzynarodowy znak towarowy, miejsce zamieszkania*, zob. (3).

(3) Przez postać elektroniczną zgłoszenia należy rozumieć postać ustaloną przy przesyłaniu zgłoszenia przy zastosowaniu sieci telekomunikacyjnej lub na informatycznym nośniku danych w rozumieniu art. 3 pkt 1 ustawy z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne.

3) [base="(ilekroć)"]

Ta formuła pozwala wyszukać LJT, w których definicjach została użyta konstrukcja *ilekroć w ... (jest) mowa o ...*. Otrzymane wyniki są zbieżne z otrzymanymi w wyniku zastosowania formuły 2), gdyż w dalszej części definicji występuje forma czasownika *rozumieć*, por. (4).

(4) Ilekroć w ustawie jest mowa o: 1) znakach towarowych – rozumie się przez to także znaki usługowe [...].

Możliwe jest także zastosowanie formuł wyszukiwawczych z innymi wyrażeniami, np. *... znaczy tyle, co, ... znaczy, ... oznacza*. W korpusie Pwp ich użycie zwróciło zbiory puste, ale formuły mogą być przydatne do identyfikacji LJP w innych tekstach prawnych.

¹⁷ Zasadniczo fraza n-składnikowa. Podaną formułę należy potraktować jako bazową. Po niewielkich modyfikacjach może zostać wykorzystana do identyfikacji jednostek ponaddwuelementowych. Miejsce znaku „[]” może zająć oznaczenie klasy gramatycznej, co pozwoli wyszukiwać frazy rzeczownikowe z podrzednikami należącymi do określonych klas gramatycznych: rzeczowników, przymiotników lub przyimków.

Do wyszukiwania LJT zdefiniowanych kontekstowo, a więc wg różnych schematów konstrukcyjnych, posłużyć się można natomiast danymi zestawionymi na listach słownictwa charakterystycznego i kolokacji. Jak wiadomo, frazy te są szeregowane na podstawie miary istotności C-value (Marciniak, Mykowiecka, Rychlik, 2017, s. 66–67), dlatego na czele listy należy się spodziewać wystąpień LJT. Na zamieszczonej w Tabeli 3. liście słownictwa charakterystycznego znalazły się głównie LJT mające definicje równościowe, z powodzeniem wydobywane z korpusu za pomocą ww. formuł wyszukiwawczych, np.: *znak towarowy, wzór przemysłowy, wzór użytkowy, oznaczenie geograficzne, topografia układu scalonego*. Definicję kontekstową mają natomiast LJT *prawo ochronne* i *dodatkowe prawo ochronne*, por. (5).

(5) Art. 95 1. Na wzór użytkowy może być udzielone prawo ochronne. 2. Przez uzyskanie prawa ochronnego nabywa się prawo wyłącznego korzystania ze wzoru użytkowego w sposób zarobkowy lub zawodowy na całym obszarze Rzeczypospolitej Polskiej.

To, że dane wyrażenie ze wstępnej listy PWJJ zostało zidentyfikowane jako LJT lub jej segment, nie wyklucza oczywiście, że w badanym tekście jest ono zarazem wykładnikiem innej JTP. Aby to sprawdzić, należy przeszukać korpus pod kątem współwystępowania segmentów zidentyfikowanych LJT z wyrażeniami o innych kształtach. Na przykład dla LJP *stan techniki* do badania można wykorzystać formuły: 1) [!base="(stan)"] [!orth="(techniki)"] oraz 2) [!base="(stan)"][!base="(technika)"]. Pierwsza zwróciła 12 przykładów, druga 5, por. *jest w stanie, stan majątkowy, stan rodzinny, stan prawny, stan zdrowia, stan zatrudnienia, stan towarów, stan zagrożenia bezpieczeństwa Państwa, stan zagrożenia ważnych interesów Państwa, dziedzina techniki, nowoczesne techniki, techniki naboru*. Badanie potwierdziło, że zamieszczone na wstępnej liście PWJJ lematy *stan, technika*, o frekwencji odpowiednio 26 i 19, nie tylko są segmentami budującymi LJT *stan techniki*, ale także są wykładnikami innych jedno- lub wieloczłonowych jednostek. Jest jasne, że im wyższa frekwencja lematów na liście PWJJ, tym większe prawdopodobieństwo tego, że dany lemat jest wykładnikiem więcej niż jednej JTP. Lematy o frekwencji powyżej 1, w szczególności zaś te o najwyższej frekwencji, jak np. rzeczowniki zamieszczone w trzeciej kolumnie Tabeli 2., wymagają szczegółowych badań obejmujących dokładną analizę wszystkich przykładów ich użycia. Zwróćmy uwagę, że rozpoczynający listę rzeczownikowy lemat *prawo* ma w korpusie Pwp frekwencję 617, a przecież pracujemy z bardzo małym korpusem. Dokładne zanalizowanie wszystkich kontekstów jego użycia, co jest konieczne, gdy chcemy rzetelnie zidentyfikować wszystkie JTP, których jest wykładnikiem lub segmentem, będzie czynnością bardzo czasochłonną. Proces ten można by znacznie przyspieszyć, gdyby Korpusomat dawał możliwość szybkiego tworzenia podkorpusu z wyszukanych przykładów, który stawałby się automatycznie nowym korpusem Korpusomatu i – jak korpus podstawowy – poddawany mógłby być analizie statystycznej i gramatycznej. Wyniki zapisane w postaci plików o formacie csv lub xls można co prawda przeszukiwać i analizować, a także można z nich utworzyć nowy korpus, ale wymaga to wykonania kilkunastu operacji.

2.2.3. SJT

Ze względu na brak definicji w danym tekście prawnym zautomatyzowanie wyszukiwania SJT jest zadaniem znacznie trudniejszym niż wyszukanie LJT. Do identyfikacji SJT można natomiast wykorzystać oferowane w aplikacji narzędzia statystyczne: wcześniej wspomniane pochodzące z aplikacji TermoPL narzędzie do wydobywania terminologii dziedzinowej oraz narzędzia do identyfikacji kolokacji (por. Tabelę 3.). Szczegółowej analizie jakościowej należy poddać głównie wyrażenia znajdujące się na pierwszych miejscach list zestawianych w Korpusomacie, które nie mają definicji legalnych. Punktem wyjścia powinno być jednak zweryfikowanie, czy wyrażenie typowane w Korpusomacie jako kolokacja, nie jest wielowyzrazową jednostką leksykalną odnotowaną w ogólnym słowniku języka polskiego lub powszechnie znaną nazwą własną. Jeśli tak, należałoby je uznać za PWJJ, a nie STJ. Taką interpretację będzie miała np. jednostka *w dobrej wierze* (zob. WSJP) i nazwy własne na liście z Korpusomatu notowane jako Urząd Patentowy, Rzeczpospolita Polska, które faktycznie trzeba interpretować jako jedną nazwę własną – Urząd Patentowy Rzeczypospolitej Polskiej¹⁸. Następnie analiza powinna przebiegać dwutorowo.

Po pierwsze należy sprawdzić, czy wyszukane w programie wyrażenia zostały zdefiniowane w kompendiach prawnych. Dla wyrażen z zakresu prawa własności przemysłowej przydatnym źródłem wiedzy nt. terminologicznego użycia danego wyrażenia jest np. słownik terminów publikowany na stronie Urzędu Patentowego RP (zob. przyp. 16). Trzeba jednak pamiętać, że terminy odnotowywane w tego rodzaju kompendiach niekoniecznie mają status pojedynczych JTP, dlatego muszą być poddawane językoznawczym procedurom identyfikowania jednostek leksykalnych, w szczególności testom substytucji potwierdzającym wzajemną dwustronną rozłączność (zob. np. Bednarek, Grochowski, 1993, s. 33–38). Nierzadko bowiem w kompendiach prawnych jako termin traktowane jest połączenie JTP. Taki status ma np. termin *badanie stanu techniki*¹⁹, składający się z dwóch JTP: *badanie* i *stan techniki*, por. *badanie stanu techniki, sprawozdanie o stanie techniki, poszukiwanie w stanie techniki, badanie zgłoszeń dotyczących wynalazków, badanie takiego zgłoszenia*. Zdarza się i tak, że pojęciu specjalistycznemu inny termin odpowiada w kompendium prawniczym, inny lub inne w tekście prawnym. W słowniku terminów publikowanym na stronie internetowej Urzędu Patentowego uwzględniony został np. termin *prawo wyłączne*:

Prawo podmiotowe, skuteczne *erga omnes* (wobec wszystkich), zapewniające podmiotowi uprawnionemu monopol prawny na korzystanie z określonego dobra prawnego (wynalazku, wzoru użytkowego, wzoru przemysłowego, znaku towarowego, oznaczenia geograficznego, topografii, topografii układu scalonego, utworu)²⁰,

¹⁸ Taki 4-gramowy termin jest zresztą wydobywany w programie TermoPl. Wynik uzyskany w Korpusomacie jest mniej dokładny.

¹⁹ Zob. <https://uprp.gov.pl/pl/slownik-terminow?litera=B> (dostęp: 1.04.2023).

²⁰ Zob. <https://uprp.gov.pl/pl/slownik-terminow?nazwa=Prawo-wy%C5%82%C4%85czne> (dostęp: 1.04.2023).

który nie został ani razu użyty w korpusie Pwp, mimo że samo pojęcie odnosi się do regulacji prawnych zawartych w ustawie „Prawo własności przemysłowej”. W korpusie natomiast wyszukiwanie na podstawie formuły [base="(wyłączny)"] zwróciło łącznie 7 wyników, z których 5 było nośnikiem treści pojęcia wyrażonego w słowniku Urzędu Patentowego terminem *prawo wyłączne*. Były to wyrażenia: *prawo wyłącznego korzystania z wynalazku*, *prawo wyłącznego korzystania ze wzoru użytkowego*, *prawo wyłącznego korzystania z wzoru przemysłowego*, *prawo do wyłącznego korzystania z topografii* oraz *prawo wyłącznego używania znaku towarowego*. Z językoznawczego punktu widzenia wymienione wyrażenia są więc konstrukcjami zawierającymi: 1) LJT: *wynalazek*, *wzór użytkowy*, *wzór przemysłowy*, *topografia*, *znak towarowy* oraz 2) SJT *prawo (do) wyłącznego korzystania z (czego) __ || używania (czego) __*.

Po drugie, należy wyszukać w korpusie wszystkie użycia badanych słów charakterystycznych lub kolokacji i – odwołując się do intuicji językowej oraz danych leksykograficznych – przeanalizować je pod kątem niestandardowej łączliwości leksykalnej i semantyczno-leksykalnej. Jak bowiem wiadomo, może być ona efektem zmiany znaczenia wyrażenia i ukonstytuowania się nowej jednostki leksykalnej. Formalnym wykładnikiem takiej zmiany nierzadko bywa obecność/brak przyimka w kontekście danego wyrażenia. Tę właściwość można z kolei wykorzystać do przeprowadzenia diagnostycznych wyszukiwań w Korpusomacie. Uzyskane wyniki mogą z kolei podpowiedzieć kierunki dalszych analiz i interpretacji. Zilustrujmy tę procedurę na przykładzie rzeczownika *prawo*.

Zidentyfikowaliśmy już jednostkę typu SJT *prawo (do) wyłącznego korzystania z (czego) __ || używania (czego) __*. Powstaje pytanie, jakie inne JTP zawierają segment *prawo* i jakiego typu są to JTP, czy jest wśród nich SJT. Zastosujmy formułę wyszukiwawczą [base="(prawo)"][pos="(prep)"] wraz z włączoną funkcją „Statystyki”, grupowanie wg lematu. Otrzymujemy 193 wyniki oraz listę wystąpień badanego lematu z 6 przyimkami, zob. Tabelę 4.

Lp.	Wyrażenie przyimkowe	Frekwencja w korpusie Pwp
1.	<i>prawo z</i>	160
2.	<i>prawo do</i>	23
3.	<i>prawo o</i>	4
4.	<i>prawo w</i>	3
5.	<i>prawo przez</i>	2
6.	<i>prawo na</i>	1

Tabela 4. Przyimki współwystępujące z lematem *prawo*

Wyrażenia *prawo przez* wyłączamy z dalszej analizy, gdyż współwystępowanie na linii tekstu nie jest skorelowane z zależnością składniową, co identyfikujemy intuicyjnie,

a potwierdzamy, interpretując związki składniowe w przykładach z tymi ciągami, por. *Patent wygasa na skutek [...] zrzeczenia się prawa przez uprawnionego*. Przyimek *przez* wchodzi w skład konstrukcji *zrzeczenia się przez uprawnionego*, a więc jest bezpośrednio zależny od formy *zrzeczenia się*, a nie *prawo*. Również ciągi *prawo o* należy zinterpretować jako nietworzące grupy syntaktycznej, gdyż przyimek nie wchodzi w związek zależności z formą *prawo*, ale z formą *powiadomienia*, por.:

(6) Roszczenia z tytułu naruszenia prawa z rejestracji wzoru przemysłowego, jeżeli sprawca działał w dobrej wierze, mogą być dochodzone za okres rozpoczynający się po dniu dokonania ogłoszenia w „Wiadomościach Urzędu Patentowego” informacji o udzieleniu prawa, a w przypadku wcześniejszego powiadomienia przez uprawnionego osoby naruszającej prawo o dokonany zgłoszeniu – od daty tego powiadomienia.

Z perspektywy składni ogólnej polszczyzny połączenia *prawo z (czego)* i *prawo na (czym)* są niestandardowe. Przykłady ich użycia w korpusie Pwp wskazują, że nie są to ciągi przypadkowe, ale konstrukcje składniowe, por.: *prawo ochronne lub prawo z rejestracji, uzyskanie prawa z rejestracji, wydanie decyzji o udzieleniu prawa z rejestracji wzoru przemysłowego, naruszenie prawa z rejestracji, prawo z rejestracji topografii może być udzielone, zgoda osób, którym służą prawa na patencie*. Konstrukcje te należy zarazem odróżnić od standardowych połączeń *prawo do (czego)*_, por.: *twórca wynalazku, wzoru użytkowego albo wzoru przemysłowego ma prawo do wynagrodzenia za korzystanie z tego wynalazku, zgłaszający, który nabył prawo do korzystania z uprzedniego pierwszeństwa przysługującego ze zgłoszenia*, w których identyfikujemy jednostkę *prawo do (czego)*_, mającą znaczenie: ‘określonego rodzaju możliwość, która przysługuje komuś zgodnie z obowiązującymi przepisami lub regułami przyjętymi w jakiejś społeczności lub wynika z następstwa zdarzeń’ (WSJP). Biorąc pod uwagę zróżnicowanie wyrażeń przyimkowych rozwijających niestandardowe użycia wyrażenia *prawo*, można postawić hipotezę, że jest ono podstawą dwóch innych jednostek: *prawo z rejestracji (czego)*_, *prawo na patencie*. Tym, co je łączy, jest specjalistyczna treść, objaśniana w kompendiach prawniczych. Wymienione wyrażenia są nazwami ochrony prawnej, która jest udzielana na określone dobro intelektualne lub materialne. Jeśli ochronie podlega topografia, wzór przemysłowy, oznaczenia geograficznego znaku towarowego ochronę tę nazywa się *prawem z rejestracji*. Jeśli ochronie podlega znak towarowy, używana jest nazwa *prawo ochronne na znak towarowy*. Wyrażenie *prawo na patencie* odnosi się do ochrony wynalazków (zob. Michniewicz, 2022; Barczewski i Kowalska (red.), 2019). Posiadanie definicji specjalistycznej pozwala wymienione jednostki potraktować jako SJT.

Należy zaznaczyć, że w trakcie identyfikacji JTP, tj. stawiania i weryfikowania hipotez o uznaniu danego wyrażenia za JTP, Korpusomat jest niezwykle przydatnym narzędziem. Pozwala bowiem szybko wyszukiwać materiał w rozmaitych konfiguracjach, tj. o różnych parametrach gramatycznych, co znacznie usprawnia żmudny proces ustalania statusu lematów wchodzących w skład wstępnej listy PWJJ.

2.2.4. SWJJ

Do tej klasy zaliczamy derywaty jednostek ogólnych, utworzone w wyniku zastosowania podstawowych mechanizmów derywacyjnych: derywacji sufiksальной przymiotników odczasownikowych (imiesłowów przymiotnikowych) i derywacji sufiksальной rzeczowników odczasownikowych (odstłowników/gerundiów). Za pomocą narzędzi Korpusomatu jesteśmy w stanie automatycznie zestawić listy lematów i przykłady ich użycia. Zastosowane formuły oraz otrzymane wyniki (przy włączonym module „Statystyki”, grupowanie wg lematu) zebrane zostały w Tabeli 6.

Klasa gramatyczna	Formuła wyszukiwawcza	Liczba przykładów	Liczba lematów czasownikowych	Przykładowe formy o najwyższej frekwencji
imiesłowy przymiotnikowe czynne	[pos="(pact)"]	560	97	zgłaszający, dotyczący, będący, orzekający, zawierający, mający
imiesłowy przymiotnikowe bierne utworzone od czasowników dokonanych	[[pos="(ppas)"&aspect="(perf)"]]	1121	co najmniej 100	określony, uchylony, uprawniony, udzielony, dokonany, scalony
imiesłowy przymiotnikowe bierne utworzone od czasowników niedokonanych	[[pos="(ppas)"&aspect="(imperf)"]]	205	56	wymagany, chroniony, używany, stosowany, rozpatrywany, prowadzony, wykorzystywany
gerundia utworzone od czasowników dokonanych	[[pos="(ger)"&aspect="(perf)"]]	1328	co najmniej 100	udzielenie, uzyskanie, zgłoszenie, unieważnienie, dokonanie, wydanie, złożenie, naruszenie
gerundia utworzone od czasowników niedokonanych	[[pos="(ger)"&aspect="(imperf)"]]	496	co najmniej 100	korzystanie, używanie, rozpatrywanie, wykonywanie, orzekanie, wprowadzanie, oferowanie

Tabela 6. Derywaty odczasownikowe w korpusie Pwp

Należy zwrócić uwagę na to, że w materiale najwięcej jest lematów związanych z bazami czasownikowymi dokonanymi. Dominacja czasowników dokonanych, a co za tym idzie także ich derywatów jest bowiem wyróżnikiem tekstów prawnych (Moroz, 2021, s. 76).

Lista wyszukanych automatycznie derywatów odczasownikowych nie jest tożsama z listą SWJJ. Wyrażenia na niej zamieszczone trzeba poddać analizie językoznawczej w celu: 1) identyfikacji jedno- i wielosegmentowych jednostek leksykalnych; 2) ustalenia reprezentowanych przez wydobyte jednostki klas JTP. Listy zestawione automatycznie zawierają bowiem pojedyncze segmenty (por. Tabelę 6.), z których część to zarazem jednostki leksykalne, czyli *signifiant* skojarzone z *signifié*, część natomiast to segmenty wchodzące w skład jednostek ponadjednosegmentowych. Identyfikacja JTP wymaga więc zanalizowania wszystkich kontekstów, w których występują formy poszczególnych lematów z ww. list.

Gdy dane wyrażenie występuje w niewielu kontekstach, wystarczy zastosować podstawową formułę wyszukiwawczą, uwzględniającą formę lematu i klasę gramatyczną derywatu. Dla gerundium *wystawienie* będzie to formuła [(base="(wystawić)"&pos="(ger)"]], która zwraca 11 przykładów, m.in.: *daty wystawienia wynalazku, wzoru użytkowego albo wzoru przemysłowego na wystawie, daty wystawienia towaru oznaczonego tym znakiem w Polsce lub za granicą, prawo do korzystania z uprzedniego pierwszeństwa przysługującego ze zgłoszenia bądź wystawienia wynalazku dokonanego przez inną osobę, dołączyć dowód potwierdzający zgłoszenie wynalazku we wskazanym państwie bądź wystawienie go na określonej wystawie*. Z ich analizy wynika, że JTP ma postać *wystawienie (czego)*. Jednostka jest jednosegmentowa i otwiera jedną pozycję składniową dla frazy nominalnej. W badanym tekście prawnym pozycja ta jest wypełniana przez frazy o określonym napełnieniu leksykalnym, co jest konsekwencją używania JTP *wystawienie* w odniesieniu do opisywanej w tekście prawnym procedury zgłaszania wynalazku lub wzoru użytkowego²¹. Znaczenie JTP nie różni się od znaczenia jednostki należącej do polszczyzny ogólnej (LJW), bo *wystawić* w WSJP jest objaśniane jako: ‘umieścić w określonym miejscu tak, aby każdy mógł przyjść i to zobaczyć oraz wykonać zwyczajowe czynności w związku z tym’. JTP *wystawienie (czego)* jest derywatem tego czasownika o tym samym znaczeniu, ale o węższym odniesieniu. Spełnia więc kryteria nałożone na JTP typu SWJJ.

Dla wyrażen o wyższej frekwencji warto czynność wiązania kształtu z treścią usprawnić, wyszukując nie tylko przykłady użycia form lematów, ale także ich kolokacje. Dla interesujących nas derywatów odczasownikowych będą to głównie połączenia z przymiotnikami, rzeczownikami i przyimkami.

Podsumowując, analiza językoznawcza listy wyszukanych automatycznie derywatów odczasownikowych musi zakończyć się identyfikacją JTP należących do poszczególnych klas jednostek. Będą to jedno- lub wielosegmentowe: 1) PWJJ, np. *doktor habilitowany*; 2) SWJJ, np. *zgłoszenie (czego)*, *wygaśnięcie (czego)*, *chroniony, korzystający z (czego)*, *uczestniczący w (czym)*, *zajmujący się (czym)*; 3) PSWJJ, np. *uprawniony z (czego)*, *zespół orzekający*; 4) SJT, np. rzeczownik *zgłaszający* (jednostka w kompendiach

²¹ Chodzi o to, że jeśli zgłaszający wykaże, że wynalazek lub wzór przemysłowy został wystawiony na wystawie, może skorzystać z uprzedniego pierwszeństwa, czyli „rezerwacji daty zgłoszenia np. patentu na wynalazek w jakimś innym miejscu niż w Urzędzie Patentowym” (Michniewicz, 2022, s. 139).

prawnych ma bowiem własną definicję, por.: ‘Strona postępowania przed Urzędem Patentowym w sprawie uzyskania patentu, prawa ochronnego albo prawa z rejestracji’²²), *pierwszeństwo z (czego)*_, por. ‘Prawo danego podmiotu (osoby fizycznej lub prawnej) do uzyskania prawa wyłącznego (monopolu) w pierwszej kolejności (przed innymi osobami uprawnionymi do ubiegania się o ochronę)’²³. Podobnie jak w korpusie tekstów z zakresu prawa karnego w korpusie Pwp najliczniejsza jest klasa SWJJ (Moroz, 2021, s. 76).

2.2.5. PSWJJ

Ciężar identyfikacji ostatniego z typów JTP spoczywa na analizie językoznawczej. Etap typowania tych jednostek może być natomiast wspomagany wyszukiwaniami przeprowadzonymi w Korpusomacie – wstępnymi (wyszukiwanie list lematów i kolokacji) oraz diagnostycznymi (wyszukiwanie przykładów użycia wyrażen wytypowanych do analizy użycia i weryfikacji statusu).

Biorąc pod uwagę wyniki analiz JTP pochodzących z tekstów prawnych reprezentujących prawo karne, PSWJJ należy szukać w zasobie rzeczowników, w tym gerundiów, i przymiotników (zob. Gębka-Wolak, Moroz, 2019, s. 126–145). Niesystemowe modyfikacje LJW, których wyszukiwanie można by wspomagać narzędziami cyfrowymi, dotyczą zaś dodania, zredukowania lub substytucji przyimka. Tym samym szczególnie przydatne do dalszych analiz będą wydobyte z korpusu Pwp listy rzeczowników, gerundiów i przymiotników z przyimkami i bez przyimków. Odwołując się do intuicji językowej, typujemy kandydatów na PSWJJ, następnie weryfikujemy intuicje, analizując wyszukane w korpusie Pwp przykłady ich użycia oraz porównując je z danymi leksykograficznymi i danymi zawartymi w kompendiach prawniczych. W rezultacie przeprowadzonych analiz proponujemy do „Słownika jednostek tekstów prawnych” jednostkę o określonym kształcie i znaczeniu.

Na podstawie wyżej zarysowanej procedury status PSWJJ proponujemy przypisać np. jednostce *dotatkowy do (czego)*_, zob. (7)–(8).

(7) Można również uzyskać patent dodatkowy do już uzyskanego patentu dodatkowego.

(8) Art. 91. Wraz z patentem głównym tracą moc patenty dodatkowe. Jeżeli patent główny traci moc z przyczyny niemającej wpływu na wynalazek będący przedmiotem patentu dodatkowego, patenty dodatkowe do patentu głównego stają się patentami i zachowują moc przez okres, na który został udzielony patent główny.

JTP nie wykazuje różnicy znaczenia w stosunku do jednostki wyjściowej *dotatkowy* (por. ‘taki, który został dołączony do pewnej całości’ (WSJP)). Ma natomiast w niestandardowy sposób zmodyfikowany kształt – jednostka została rozwinięta przez segment *do*. Zwróćmy uwagę, że ciąg *dotatkowy* w badanym tekście prawnym jest

²² Definicja pochodzi ze słownika terminów UP, zob. <https://uprp.gov.pl/pl/sownik-terminow?litera=Z> (dostęp: 1.04.2023).

²³ Zob. <https://uprp.gov.pl/pl/sownik-terminow?litera=P> (dostęp: 1.04.2023).

także segmentem SJT *patent dodatkowy*, jednostki o znaczeniu ‘patent na ulepszenie lub uzupełnienie wynalazku, które posiada cechy wynalazku, a nie może być stosowany samoistnie’²⁴.

Z kolei w wyrażeniu *zgłoszenie o (co)* mamy do czynienia z wymianą segmentu przyimkowego i z modyfikacją znaczenia LJW, zob. (9).

(9) Zgłoszenia o udzielenie patentu na wynalazek tajny lub prawa ochronnego na wzór użytkowy tajny, dokonane przed dniem wejścia w życie ustawy i nierozpatrzone do tego czasu, uważa się za zgłoszenia dokonane w celu zastrzeżenia pierwszeństwa.

W polszczyźnie ogólnej jest bowiem używana jednostka *zgłoszenie o (czym)* – ‘informacja o zdarzeniu wymagającym interwencji przekazywana odpowiednim służbom’ (WSJP), np. *zgłoszenie o awarii*, *zgłoszenie o napadzie*, *zgłoszenie o kradzieży*. W badanym korpusie są także przykłady użycia ciągu *zgłoszenie* bez przyimka, bezpośrednio z frazą dopełniaczą, tj. *zgłoszenie wzoru zdobniczego*, *zgłoszenie wzoru przemysłowego*, które ze względu na brak różnicy znaczenia w stosunku do użyć *zgłoszenie o udzielenie patentu*, proponujemy zinterpretować jako wariant pozycyjny tej samej jednostki – *zgłoszenie (czego)* || *o (co)*. W pozycji poprzyimkowej jest realizowana rzeczownikowa nazwa czynności.

Identyfikacja PSWJJ jest zadaniem trudnym i żmudnym, przypominającym szukanie igły w stogu siana. Wymaga przejrzenia nieraz bardzo wielu przykładów użycia wyrażen podejrzewanych o zmodyfikowane LJW oraz porównywania schematów składniowych realizowanych w tekście prawnym ze schematami odnotowywanymi w opracowaniach leksykograficznych.

3. Wnioski

Wnioski z przeprowadzonego badania pilotażowego zostały przedstawione niżej, z podziałem na dotyczące zastosowanego narzędzia oraz koncepcji JTP i procedury badawczej.

3.1. Zastosowane narzędzie

Ogólna ocena przydatności Korpusomatu do badań tekstów prawnych i wspomaganie opracowywania „Słownika jednostek tekstu prawnego” jest bardzo pozytywna. Atutem aplikacji jest dostęp z jednego miejsca do różnego rodzaju narzędzi analitycznych, a co za tym idzie, możliwość łatwego i szybkiego wydobycia wielu informacji o ogólnych i szczegółowych właściwościach gramatycznych i statystycznych badanego materiału. Z kolei możliwość interpretacji statystycznej danych wyszukiwanych za pomocą języka zapytań, czego efektem są listy frekwencyjne wyrażen grupowanych wg kilku kryteriów (np. lemat, część mowy) pozwala korelować informacje z zakresu morfoskładni

²⁴ Zob. <https://uprp.gov.pl/pl/slownik-terminow?litera=P> (dostęp 1.04.2023).

z danymi o zasobie leksykalnym. Przydać się może także do ograniczania materiału poddawanego szczegółowej analizie językoznawczej (koncentrujemy się na wyrażeniach najczęstszych).

Jako że język zapytań wymaga zapamiętania wielu oznaczeń, bardzo dobrym rozwiązaniem jest udostępnienie „Konstruktora zapytań”. Nie tylko przypomina elementy języka zapytań, ale także inspiruje do stosowania różnych ich kombinacji.

Korpusomat korzysta z kilku narzędzi zewnętrznych, w tym z analizatora morfologicznego Morfeusz i tagera Concraft. Słabą stroną przeprowadzanego tagowania okazała się identyfikacja partykuł, wykrzykników, rozkaźników i predykatywów.

Z kolei zestaw wyrażen zaproponowany w oknie „Słownictwo charakterystyczne” był bardzo trafny.

Nie udało się wykorzystać narzędzia do porównywania korpusów. Po kilkunastu godzinach od uruchomienia narzędzie nie skończyło pracy (brak informacji o postępie przetwarzania), mimo że korpus Pwp był porównywany z równie małym korpusem. Próbę porównania podjęto kilka razy, zawsze bez rezultatu.

Podczas pilotażu nie przetestowano natomiast narzędzia „Profile słów”, gdyż informacje uzyskiwane o funkcjach składniowych i związkach zależności nie wydawały się przydatne do identyfikacji typów JTP. Nie jest wykluczone, że narzędzie przyda się na kolejnych etapach tworzenia „Słownika jednostek tekstu prawnego”.

Jeśli Korpusomat będzie rozwijany, można pomyśleć o dodaniu możliwości szybkiego (jeden przycisk) tworzenia podkorpusu z wyszukanych przykładów, który stawałby się automatycznie nowym korpusem Korpusomatu i – jak korpus podstawowy – poddawany mógłby być analizie statystycznej i gramatycznej. Przydałaby się również możliwość gromadzenia w jednym miejscu wybranych wyników danego wyszukiwania lub kolejnych, czyli przygotowania zbioru przykładów, które chcemy dokładniej zanalizować.

3.2. Koncepcja JTP i procedura badawcza

Zastosowanie aparatu JTP do badania leksyki tekstów prawnych pozwala na nowo zinterpretować wybrane zagadnienia dotyczące leksykalnej, a z racji przyjętej metodologii częściowo także składniowej warstwy tekstów prawnych. Umożliwia opis leksyki zarówno z perspektywy jej przynależności do systemu języka etnicznego, jak i z perspektywy działań ją kształtujących, wynikających z realizacji specjalistycznej funkcji, tj. dostosowywania jej do wyrażania treści prawnych. Badania przeprowadzone na materiale reprezentującym gałąź prawa o innym charakterze i przedmiocie regulacji niż wcześniej badane prawo karne zweryfikowało skuteczność opisu warstwy leksykalnej wybranego tekstu prawnego za pomocą aparatu JTP. Przyjęty zasób klas obiektów leksykalnych okazał się wystarczający do opisu materiału leksykalnego. Okazało się natomiast, że w ramach koncepcji JTP nie zostało uwzględnione zagadnienie nazw własnych, które przecież są używane w tekstach prawnych i które z powodzeniem są automatycznie wydobywane przez aplikacje do tego celu przeznaczone. Kwestia statusu nazw własnych w koncepcji JTP powinna być więc poddana dyskusji.

Jeśli chodzi o stosowane procedury badawcze, przeprowadzone badanie potwierdziło słuszność decyzji o poddawaniu analizie dziedzinowych korpusów prawnych, czyli inaczej niż w badaniach A. Malinowskiego, które zostały oparte na dużym korpusie, składającym się z próbek bardzo wielu tekstów prawnych. Identyfikacja poszczególnych klas JTP nie może być przeprowadzona automatycznie, analizy językoznawcze są konieczne, szczególnie w zakresie identyfikacji SWJJ i PSWJJ. Im mniejszy i tematycznie bardziej spójny materiał, tym łatwiej przeprowadzić takie analizy. Narzędzia do ekstrakcji terminologii i kolokacji podpowiadają bowiem wyrażenia specyficzne dla danej gałęzi prawa. Identyfikując LJT i SJT, można zaś posiłkować się dziedzinowymi kompendiami prawniczymi. Poza tym łatwiej zauważyć błędy automatycznego tagowania morfoskładniowego.

Obiecujące wydaje się także zastosowanie procedury falsyfikacyjnej, tj. przyjęcie, że punktem wyjścia badań jest automatycznie wydobyta z materiału lista PWJJ, których status następnie jest weryfikowany. Przy czym kolejność identyfikacji JTP powinna przechodzić od LJT i SJT, przez PSWJJ do SWJJ.

Bibliografia

- Barczewski, M., Kowalska, E. (red.). (2019). *Leksykon prawa własności intelektualnej. 100 podstawowych pojęć*. Warszawa: Wydawnictwo C.H. Beck.
- Bednarek, A., Grochowski, M. (1993). *Zadania z semantyki językoznawczej*. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika.
- Bogusławski, A. (1988). Preliminaria gramatyki operacyjnej. *Polonica*, 13, 163–223.
- Bogusławski, A. (2008). *Semantyka, pragmatyka. Leksykografa głos demarkacyjny*. Warszawa: Wydawnictwo TAKT.
- Bogusławski, A., Danielewiczowa, M. (2005). *Verba Polona Abscondita. Sonda słownikowa III*. Warszawa: Elma Books.
- Bogusławski, A., Wawrzyńczyk, J. (1993). *Polszczyzna, jaką znamy. Nowa sonda słownikowa*. Warszawa: Uniwersytet Warszawski, Katedra Lingwistyki Formalnej.
- Danielewiczowa, M. (2012). *W głąb specjalizacji znaczeń. Przysłówkowe metapredykaty atestacyjne*. Warszawa: BEL Studio.
- Danielewiczowa, M. (2014). *Struktura i znaczenie eksklamatywów. Na materiale współczesnej polszczyzny*. Warszawa: Uniwersytet Warszawski, Katedra Lingwistyki Formalnej.
- Danielewiczowa, M. (2016). *Dosięgnąć przedmiotu. Rzecz o Ferdynandzie de Saussurze*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Danielewiczowa, M. (2021). *Aspekt tematyczny w informacyjnej strukturze wypowiedzi. Rozszerzanie i integracja wiedzy*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Dobaczewski, A. (2002). *Zjawiska percepcji wzrokowej. Studium semantyczne*. Warszawa: Uniwersytet Warszawski, Katedra Lingwistyki Formalnej.
- Dobaczewski, A. (2009). Operacje iterujące w języku polskim (wprowadzenie do opisu). *Poradnik Językowy* 9, 26–36.
- Dobaczewski, A. (2018). *Powtórzenie jako zjawisko tekstowe i systemowe. Repetycje, reduplikacje i quasi-tautologie w języku polskim*. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

- Dobaczewski, A., Sobotka, P., Żurowski, S. (2018). *Słownik reduplikacji i powtórzeń polskich. Od zleksykalizowanych podwojeń do regularnych układów repetycyjnych*. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.
- Gębka-Wolak, M., Moroz, A. (2019). *Jednostka tekstu prawnego w ujęciu teoretycznym i praktycznym*. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.
- Grochowski, M. (1982). *Zarys leksykologii i leksykografii*. Toruń: Uniwersytet Mikołaja Kopernika.
- Grochowski, M. (1983). Metatekstowa interpretacja parentezy. W: T. Dobrzyńska i E. Janus (red.), *Tekst i zdanie* (s. 247–258). Wrocław: Zakład Narodowy im. Ossolińskich.
- Grochowski, M. (1987). O kryteriach określania przedmiotu leksykologii. *Studia Linguistica Polono-Jugoslavica*, 5, 151–155.
- Grochowski, M. (1993). *Konwencje semantyczne a definiowanie wyrażen językowych*. Warszawa: Zakład Semiotyki Logicznej Uniwersytetu Warszawskiego.
- Grochowski, M. (1997). *Wyrażenia funkcyjne. Studium leksykograficzne*. Kraków: Wydawnictwo Instytutu Języka Polskiego PAN.
- Grochowski, M. (2001). Nieciągłe jednostki języka z segmentem nie. Wybrane problemy składni zewnętrznej, wewnętrznej i leksykografii. W: M. Balowski i W. Chlebda (red.), *Frazeografia słowiańska* (s. 215–220). Opole: Wydawnictwo Uniwersytetu Opolskiego.
- Grochowski, M. (2002). Wielowyzrazowe jednostki funkcyjne. Wprowadzenie do problematyki. W: A.M. Lewicki (red.), *Problemy frazeologii europejskiej V* (s. 43–50). Lublin: Norbertinum.
- Grochowski, M. (2012). Słownik i gramatyka – centrum języka i językoznawstwa. Wprowadzenie do problematyki. W: S. Gajda (red.), *Horyzonty humanistyki* (s. 85–104). Opole: Wydawnictwo Uniwersytetu Opolskiego.
- Grochowski, M. (2017). Terminologia a jednostki leksykalne systemu ogólnego. O definiowaniu terminów w ustawach o nauce i szkolnictwie wyższym. W: R. Przybylska i W. Śliwiński (red.), *Terminologia specjalistyczna w teorii i praktyce językoznawców słowiańskich* (s. 11–28). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- ISJP – Bańko, M. (red.). (2000). *Inny słownik języka polskiego* (t. 1–2). Warszawa: Wydawnictwo Naukowe PWN.
- Kieraś, W., Kobyliński, Ł. (2021). Korpusomat – stan obecny i przyszłość projektu. *Język Polski*, 2, 49–58.
- Malinowski, A. (2006). *Polski język prawny. Wybrane zagadnienia*. Warszawa: Wydawnictwo Lexis Nexis.
- Malinowski, A. (2012). *Polski tekst prawny*. Warszawa: Wolters Kluwer Polska.
- Marciniak, M., Mykowiecka, A., Rychlik, P. (2017). Automatyczne wydobywanie terminologii dziedzinowej z korpusów tekstowych. *Język Polski*, 1, 64–74.
- Michniewicz, G. (2022). *Ochrona własności intelektualnej*. Warszawa: C. H. Beck.
- Moroz, A. (2021). Gerundialne jednostki tekstu prawnego – problem identyfikacji. *Poznańskie Studia Polonistyczne. Seria Językoznawcza*, 1, 71–86.
- Muras, Z. (2020). *Podstawy prawa*. Warszawa: C. H. Beck.
- Przepiórkowski, A. i in. (red.). (2012). *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- WSJP – Żmigrodzki, P. (red.). (od 2007). *Wielki słownik języka polskiego PAN*: Instytut Języka Polskiego PAN. <https://wsjp.pl/> (dostęp: 1.04.2023).
- Zieliński, A. (1999). *Języki prawne i prawnicze*. W: W. Pisarek (red.), *Polszczyzna 2000. Orędzie o stanie języka na przełomie tysiącleci* (s. 50–74). Warszawa: Ośrodek Badań Prasoznawczych UJ.
- Zieliński, A. (2004). *Język prawny, język administracyjny, język urzędowy*. W: E. Malinowska (red.), *Język – prawo – społeczeństwo* (s. 9–18). Opole: Wydawnictwo Uniwersytetu Opolskiego.

STRESZCZENIE

Słowa kluczowe: język prawny, jednostka tekstu prawnego, językoznawstwo korpusowe, Korpusomat, prawo własności przemysłowej

W artykule omówiono badanie, którego celem było sprawdzenie przydatności aplikacji webowej Korpusomat do analizy warstwy leksykalnej tekstu prawnego. Ramę metodologiczną badania stanowiła koncepcja jednostki tekstu prawnego, odwołująca się do teorii jednostki języka i gramatyki operacyjnej A. Bogusławskiego. Materiałem do badań była ustawa „Prawo własności przemysłowej”, zapisana w postaci cyfrowego korpusu. Ustalono, które narzędzia aplikacji Korpusomat i w jakim zakresie można wykorzystać do wydobycia informacji o właściwościach gramatycznych i statystycznych badanego tekstu prawnego oraz identyfikacji występujących w nim pięciu typów specjalistycznych jednostek leksykalnych. Opracowano zasady badania tekstu prawnego łączące analizę wspomaganą narzędziami cyfrowymi z analizą językoznawczą. Uzyskane wyniki w przyszłości zostaną wykorzystane do opracowania „Słownika jednostek tekstu prawnego”.

SUMMARY**Digital corpora and digital tools in linguistic analyses of legal problems. Legal text units in Industrial Property Law**

Keywords: legal language, legal text unit, corpus linguistics, Korpusomat application, Industrial Property Law

The article discusses the results of a study aimed at verifying how and to what extent the Internet application Korpusomat may be used for analyzing the lexical layer of legal texts. The study was rooted in the concept of legal text unit, which, in turn, relates to A. Bogusławski's operational grammar and language units theory. The analysis was based on the Industrial Property Law, which is available in text form as a digital corpus. We describe the applicability of different Korpusomat tools in the process of extracting grammatical and statistic data concerning Industrial Property Law, and in identifying five types of specific language units which appear in the text. We establish ground rules for analyzing legal texts which combine traditional linguistic analysis and digital tools-based analysis. The results will later be used to develop a “Dictionary of Legal Text Units”.