

Monika Zaśko-Zielińska (<https://orcid.org/0000-0001-6333-660X>)  
Uniwersytet Wrocławski

## Odniesienia do korpusów w treści orzeczeń polskich sądów

W ostatnich latach ukazały się dwa artykuły Zygmunta Tobora i Mateusza Zeiferta na temat możliwości wykorzystania analizy korpusowej do ustalania znaczenia w ramach wykładni prawa (2020; 2021). Jako że obydwaj autorzy reprezentują nauki prawne, a zagadnienie dotyczy metodologii językoznawczej, sądzę, że warto kontynuować i rozszerzać podjętą dyskusję na temat korpusów w lingwistycznych analizach problemów prawnych, aby budować pomosty między językoznawstwem a prawem (Zaśko-Zielińska, 2021). Przed publikacją przywołanych na wstępie artykułów tylko wyjątkowo nawiązywano do „wykorzystania korpusów w polskiej nauce prawa” (Tobor, Zeifert, 2021), chociaż literatura naukowa na świecie obfituje w teksty podejmujące niniejszą tematykę, co zaowocowało badaniami prowadzonymi pod nazwą *legal corpus linguistics* (Bernstein, 2021) czy *corpus linguistics and the law* (Gries, 2021)<sup>1</sup>. Jednocześnie wydaje się, że eksplozja badań na styku lingwistyki korpusowej i nauk prawnych, o której pisze Stanisław Goźdz-Roszkowski (2021), wciąż nie przekłada się na praktykę polskich sądów. Przykładowo od momentu zakończenia prac nad Polskim Korpusem Listów Pożegnalnych (Zaśko-Zielińska, 2013) z jego zasobów skorzystało wielu badaczy, ale wśród użytkowników nie było ani jednego praktyka.

Warto przypomnieć, że przez ostatnie lata systematycznie pisano o wykorzystaniu słowników na potrzeby wykonywania prawa<sup>2</sup>, więc zapewne podobnego zaangażowania wymaga informowanie o możliwościach zastosowania korpusów, gdyż są one podstawą tworzenia wszystkich najnowszych słowników ogólnych, począwszy od wydanego w 2000 roku „Innego słownika języka polskiego PWN” (ISJP, 2000).

Celem niniejszego artykułu jest sprawdzenie: w jaki sposób w orzeczeniach polskich sądów przytaczana jest informacja o źródle korpusowym; jakie problemy relacjonowane w orzeczeniach były rozwiązywane przy użyciu korpusów; oraz jakie komentarze na temat wykorzystania metodologii korpusowej pojawiły się w analizowanych dokumentach.

### Material

Zestawiony do analizy materiał stanowi treść orzeczeń polskich sądów z lat 2012–2019, w których pojawiła się informacja o wykorzystaniu *korpusu* jako

---

<sup>1</sup> Por. *language and law obok forensic linguistics*.

<sup>2</sup> Por. m. in.: (Choduń, Zieliński, 2008; Bielska-Brodziak, Tobor, Żmigrodzki, 2008; Żurowski, 2014; Czelakowska, Kubicka, Klubińska, 2016; Suska, 2019).

elektronicznego zasobu tekstów języka polskiego. Orzeczenia zostały pobrane z bazy „Legalis”<sup>3</sup>, w której możliwe jest wyszukiwanie przez słowo kluczowe i przegląd orzecznictwa na dany temat<sup>4</sup>. Orzeczenia Sądu Najwyższego umieszczone w bazie „Legalis” zostały wydane po roku 1989. Natomiast prace nad Korpusem PWN rozpoczęły się w 1997 roku, a nad projektem NKJP w 2007 (Przepiórkowski i in., 2012, 7–8). Zatem zakres czasowy orzeczeń w bazie „Legalis” jest wystarczający do wyszukania odniesień na temat użycia korpusu językowego przez biegłych. Zestawienie znalezionych materiałów znajduje się w tabeli nr 1, która zawiera oznaczenia wyroków stosowane w niniejszej publikacji oraz pełne tytuły orzeczeń. Uwzględnione dokumenty obejmują tylko te postępowania, w których orzeczenia zawierały wyraz *korpus* w przywołanym powyżej znaczeniu. Można założyć, że w wielu opiniach biegłych językoznawców pojawiały się odniesienia do korpusów, ale nie zawsze ślady tych poszukiwań są widoczne w orzeczeniach sądów<sup>5</sup>, gdyż zapewne w konkretnych sprawach uznano, że nie była to informacja kluczowa w podejmowaniu decyzji przez sąd. Celem artykułu nie jest zatem ilościowa ocena odniesień do korpusów w analizowanych dokumentach, ale sposób przywoływania źródeł korpusowych w orzeczeniach, informowanie o sposobie pobierania danych oraz reprezentatywność korpusu w kontekście celu jego wykorzystania.

Lp.	Tytuł orzeczenia	Oznaczenie używane w artykule
1.	Wyrok Wojewódzkiego Sądu Administracyjnego siedzibą w Warszawie z dnia 9 stycznia 2012 r.	Mixelko1/WSA/2012
2.	Wyrok Naczelnego Sądu Administracyjnego z dnia 16 października 2013 r.	Mixelko2/NSA/2013
3.	Wyrok Sądu Okręgowego w Łodzi, X Wydział Gospodarczy, z dnia 18 lutego 2014 r.	Prenatal1/SO/2014
4.	Wyrok Sądu Apelacyjnego w Łodzi, I Wydział Cywilny z dnia 13 listopada 2014 r.	Prenatal2/SA/2014
5.	Wyrok Sądu Apelacyjnego w Białymstoku – I Wydział Cywilny z dnia 26 czerwca 2015 r.	Zabójca1/SA/2015
6.	Wyrok Sądu Okręgowego w Warszawie, X Wydział Karny Odwoławczy, z dnia 22 sierpnia 2018 r.	Zniesławienie1/SO/2018
7.	Wyrok Wojewódzkiego Sądu Administracyjnego, siedziba w Warszawie, z dnia 19 marca 2019 r.	Trybunalskie1/WSA/2019
8.	Wyrok Naczelnego Sądu Administracyjnego z dnia 30 maja 2019 r.	Cinkciarz1/NSA/2019

<sup>3</sup> <https://www.beck.pl/system-informacji-prawnej-legalis/>

<sup>4</sup> Por. [https://gov.legalis.pl/wp-content/uploads/2018/10/Instrukcja\\_korzystania\\_z\\_Systemu\\_Legalis\\_Administracja.pdf](https://gov.legalis.pl/wp-content/uploads/2018/10/Instrukcja_korzystania_z_Systemu_Legalis_Administracja.pdf)

<sup>5</sup> Por. (Piwowar, 2019).

## Orzeczenie sądu – struktura tekstu

Analizowane orzeczenia sądów to wyroki zawierające wydane przez sądy decyzje, w których znajdują się rozstrzygnięcia poszczególnych spraw. Maria Teresa Lizisowa, charakteryzując gatunki tekstów komunikujących prawo, umieściła orzeczenie sądowe wśród pisanych tekstów prawniczych, które są opracowane administracyjnie; należą do intertekstów, gdyż „w nawiązaniach międzytekstowych naśladują w sposób mniej lub bardziej oczywisty i jednoznaczny wyspecjalizowane struktury tekstu prawnego i objaśniają ich sens” (Lizisowa, 2018, s. 94–95). Publikacje odnoszące się do genologii orzeczeń koncentrują się na różnych ich wymiarach, np.: miejscu orzeczenia w dyskursie sądowym (Gizbert-Studnicki, 2007), podobieństwach między gatunkami utrwalonymi w różnych językach (Nowak-Korcz, 2020, s. 33–47), strategii formułowania uzasadnień (Rzucidło-Grochowska, 2017, s. 59–72) czy leksykalnych wykładników kategorii słuszności (Podoliński, 2018, s. 181–195). Natomiast w kontekście podjętego zagadnienia istotny jest szczególnie intertekstualny charakter orzeczenia sądowego, który znajduje odzwierciedlenie w konstrukcji tego gatunku.

Struktura omawianego tekstu obejmuje ramę z powtarzającymi się elementami: nagłówkiem identyfikujący wyrok, informacją o składzie sądu, sentencją i uzasadnieniem z ostateczną decyzją. Zawarte w orzeczeniu uzasadnienie stanowi najbardziej rozbudowaną część tekstu, gdyż obejmuje zarówno fragmenty referujące to, co ustaliły sądy niższej instancji (sygnalizowane np. zapowiedzią: *Do wydania powyższej decyzji doszło w następującym stanie faktycznym i prawnym* – Mixełko1/WSA/2012) – tzw. część historyczna, dalej – informację o zaskarżeniu wyroku, jak i ocenę sądu, który jako ostatni zajmował się sprawą (część poprzedzona zwyczajową formułą: *... Sąd zważył, co następuje*<sup>6</sup>) – tzw. część prawna. W związku z tym, jeśli sprawa przechodziła przez kilka instancji, to w kolejnych orzeczeniach mogą się powtarzać pewne fragmenty (często w różny sposób zredagowane), które przywołane zostały po to, aby w dalszej części tekstu można się było do nich odnieść. Mimo wielogłosowości tekstu orzeczenia o jego ostatecznym kształcie decyduje ostatni nadawca, przywołując istotne w przygotowaniu orzeczenia etapy postępowania oraz dokonując ich ponownej oceny. Należy uznać, że sposób redagowania tekstu jest warunkowany dwoma czynnikami: „realizacją obowiązku proceduralnego nałożonego na sąd” z jednej strony, z drugiej zaś – społecznie i kulturowo uwarunkowaną potrzebą wskazania motywów decyzji o charakterze władczym podejmowanych wobec innych podmiotów” (Rzucidło-Grochowska, 2017, s. 59)<sup>7</sup>.

W analizowanych orzeczeniach informacja o wykorzystaniu korpusu powiązana jest z historią konkretnej sprawy i wywodzi się z opinii biegłych, wypowiedzi

<sup>6</sup> Por. Mixełko1/WSA/2012; Mixełko2/NSA/2013; Prenatal1/SO/2014; Prenatal2/SA/2014; Trybunalskie1/WSA/2019; Cinkciarz1/NSA/2019; Zabójca1/SA/2015; Zniesławienie1/SO/2018.

<sup>7</sup> Por. też: Raport „Publiczna dostępność orzeczeń sądowych” przygotowany w 2010 roku przez Helsińską Fundację Praw Człowieka.

wnioskodawców, konkretnych organów oraz ich oceny w toku poprzednich postępowań. Sposób przywoływania wykorzystanego korpusu zależy zatem, z jednej strony od tego, jak wywiązywali się z tego zadania wszyscy nadawcy tekstu, gdyż orzeczenie musi dokładnie relacjonować ich działania, z drugiej strony powinien być także warunkowany uwiarygodnieniem słuszności rozumowania sądu.

### Sposób przywołania korpusu w orzeczeniu

Wśród wykorzystywanych współcześnie źródeł leksykograficznych odnoszących się do polszczyzny ogólnej znajdziemy słowniki tradycyjne (papierowe), słowniki elektroniczne, korpusy oraz zasoby internetowe traktowane czasami jako korpus. W orzeczeniach sądowych korpusy traktowane są właśnie jako jeden z zasobów leksykograficznych:

- „Uprawniony przecząc twierdzeniom wnioskodawcy podniósł natomiast, że nazwa (...) nie pojawia się w żadnym z ogólnych zbiorów leksykalnych współczesnego języka i nie jest odnotowywane przez korpus języka polskiego.” (Mixelko2/NSA/2013).
- „nierozpatrzenie przez Sąd w sposób wyczerpujący zgromadzonego materiału dowodowego, w szczególności całkowitego pominięcia przedłożonych dowodów na okoliczność wejścia oznaczenia „cinkciarz” do języka potocznego, w tym wszystkim dowodów ze słowników, Korpusu Języka Polskiego (...)” (Cinkciarz1/NSA/2019).

Punktem wyjścia do analizy zebranych orzeczeń było sprawdzenie sposobu informowania o źródle korpusowym, które zostało wykorzystane w procedurze orzeczniczej. Agnieszka Bielska-Brodziak, Zygmunt Tobor i Piotr Żmigrodzki w swojej publikacji na temat zasad używania słowników przez prawników (2008, s. 90–91) zwracają uwagę na konieczność podawania danych bibliograficznych źródeł leksykograficznych. Zasada ta dotyczy, oczywiście, również zasobów elektronicznych, chociaż zasób niezbędnych informacji nieco się zmienia. Z perspektywy lingwistycznej najważniejszą zaletą słowników elektronicznych jest możliwość ich bieżącego uzupełniania i aktualizowania na podstawie najnowszych danych językowych. Taką zaletę mają też wszystkie „korpusy monitorujące”. Z perspektywy prawa czynnikiem decydującym o wartości źródła jest jego stabilność. Stąd też informacje zaczerpnięte ze słownika elektronicznego powinny zawierać dane chronologiczne (Czelakowska, 2016, s. 21–22), np.: informację o ostatniej aktualizacji, datę najstarszego poświadczenia wyrazu, a dane korpusowe opatrywane są np. datą pobrania i potwierdzone wydrukiem.

Wśród informacji niezbędnych do zweryfikowania danych korpusowych należy ostatecznie wymienić:

- nazwę korpusu (w tym jego typ i informację o miejscu jego udostępnienia);
- informację zasadach pobierania danych.

## Identyfikacja źródła korpusowego

Mimo że przywołanie nazwy źródła leksykograficznego wydaje się najbardziej oczywistym wymogiem, nie zawsze jest ono realizowane z pełną starannością. Poszukiwania odniesień do słowników elektronicznych w orzeczeniach sądowych, które relacjonuje Marek Suska, pokazują, że autorzy tych dokumentów nie zawsze uwzględniają wspomnianą zasadę (2019, s. 58), co utrudnia identyfikację źródła leksykograficznego. Podobne wnioski należy sformułować co do możliwości rozpoznania konkretnego korpusu na podstawie analizowanych przez mnie orzeczeń. Oczywiście, liczba źródeł korpusowych dostępnych dla języka polskiego nie jest tak duża jak materiałów słownikowych, ale nowe zasoby korpusowe (tzw. małe lub tzw. duże korpusy) wciąż powstają, a czas trwania postępowań sądowych w kolejnych instancjach może trwać kilka lat, więc istnieje prawdopodobieństwo pojawiania się w tym czasie kolejnych, nowych korpusów lub ich wersji.

W analizowanych orzeczeniach źródło korpusowe opisywane było najczęściej jako *korpus języka polskiego* (czasem w wersji: tzw. *korpus języka polskiego*, w liczbie mnogiej: *korpusy języka polskiego*<sup>8</sup>). Spośród nazw własnych odnotowano: *Korpus Języka Polskiego*, *Korpus Języka Polskiego PWN*, *Narodowy Korpus Języka Polskiego* (w tym: *korpus narodowy*), *Korpus Języka Polskiego (wspólna inicjatywa polskiej Akademii Nauk oraz...)* oraz *korpusy narodowe* – w znaczeniu korpusy narodowe dla języków innych niż polski.

Ze względu na wyraźną dwuczęściową strukturę tekstów orzeczeń, która powiązana jest też z udziałem różnych nadawców, sprawdziłam również sposób odnotowywania użytych nazw korpusów w częściach referujących orzeczeń i w ich częściach zasadniczych. Niestety, mimo administracyjnego, sformalizowanego charakteru dokumentu w analizowanym materiale pojawiła się pewna dowolność w przypisywaniu nazw korpusom, co wynika być może częściowo z braku świadomości w tym zakresie, a częściowo jest efektem niestaranności redakcyjnej:

a) fragment referujący (Mixełko2/NSA/2013):

Uprawniony przeczając twierdzeniom wnioskodawcy podniósł natomiast, że nazwa (...) nie pojawia się w żadnym z ogólnych zbiorów leksykalnych współczesnego języka i nie jest odnotowywane przez korpus języka polskiego.

b) fragment ostatecznej oceny sądu (Mixełko2/NSA/2013):

Uprawniony przeczając twierdzeniom wnioskodawcy podniósł natomiast, że nazwa(...) nie pojawia się w żadnym z ogólnych zbiorów leksykalnych współczesnego języka i nie jest odnotowywane przez korpus języka polskiego, w tym przez Korpus Języka Polskiego PWN.

---

<sup>8</sup> Por. „Ww. aneks został oparty także na materiałach pozyskanych z *korpusów języka polskiego* w dniu 12 lutego 2017.” – Trybunalskie1/WSA/2019.

Aby możliwe było zweryfikowanie danych pobranych z korpusu, potrzebna jest jego nazwa, gdyż samo wyrażenie *korpus języka polskiego* mówi właściwie tylko tyle, że korpus ten zawiera teksty w języku polskim. Nazwie powinna też towarzyszyć informacja o miejscu udostępnienia korpusu, czyli adres internetowy korpusu, który pozwala na sprawdzenie wykorzystanych w orzeczeniu informacji. Zazwyczaj strona, na której umieszczone są informacje o dostępie do korpusu, zawiera także inne wartościowe dane, które umożliwiają weryfikację jakości źródła: adresy publikacji naukowych na temat korpusu czy instrukcje korzystania ze źródła.

Ze względu na to, że nazwa korpusu może zawierać różne informacje, które nie zawsze w pełni są powtarzane, osobno uwzględniłam typ korpusu, czyli np.: wskazanie jego rozmiaru; języka; zakresu (ogólny – specjalistyczny); współczesności (synchroniczny) lub historyczności językowej (diachroniczny); stanu zamkniętego (statyczny) lub monitorującego charakteru korpusu; jego zrównoważenia lub oportunistyczności. Nie zawsze każda informacja jest niezbędna, ale często konieczność jej uwzględnienia zależy od problemu rozwiązywanego w konkretnym orzeczeniu sądowym. Przykładowo w jednym z dokumentów pojawiło się uzasadnienie wykorzystania korpusów dla języków innych niż polski:

W tym zakresie należy wskazać, że biegły wyjaśnił, z jakich przyczyn analizy wyrazu „(...)” dokonał w oparciu o określone korpusy narodowe – oparł się on na najbardziej znanych językach wśród Polaków (k. 904). To strony sporu zgodnie wniosły o wydanie opinii w oparciu o określone języki (Prenatal1/SO/2014).

Natomiast w innym orzeczeniu konieczne było potwierdzenie właściwego doboru materiału i zrównoważenia przywołanego korpusu:

Skarżący nie podważył skutecznie wniosków opinii biegłej. Sam fakt, że materiał porównawczy do jej stworzenia pochodził z książek s-f, literatury kobiecej, podręcznika o tarocie, czy też wycinków z takich tytułów jak „super ekspres”, nie może decydować o negatywnej ocenie opinii biegłej. Biegła wyjaśniła, iż po podstawie sporządzania opinii z zakresu językoznawstwa jest tzw. „korpus języka polskiego”. Jest to zrównoważony tematycznie i gatunkowo zbiór tekstów polskich służących do zilustrowania użycia słów i ich znaczeń (Zabójca1/SA/2015).

Zamieszczenie powyższego wyjaśnienia w części prawnej orzeczenia było odpowiedzią na zawarte w apelacji zaskarżenie wyroku Sądu I instancji.

### **Zasady pobierania danych z korpusu**

Ze względu na to, że sposób pozyskiwania danych korpusowych wpływa na uzyskane wyniki, procedura analizy powinna być zawsze dokładnie przedstawiona, aby przy tych samych ustawieniach można było otrzymać powtarzalne rezultaty z publicznie dostępnego korpusu.

Ułatwieniem zachowania zaplanowanych ustawień jest możliwość generowania skompresowanych odsyłaczy do wyników, o czym informują autorzy wyszukiwarki w opracowaniu na temat NKJP (Przepiórkowski, 2012, s. 254), a także pobieranie wyników w formie arkusza kalkulacyjnego. Inna możliwość to używanie tej samej wyszukiwarki (Przepiórkowski, 2012, s. 257–258) oraz stosowanie powtarzalnej składni zapytań (Hebal-Jeziarska, 2013, s. 23–25), która uwzględnia np. zakres wyszukiwania w całym korpusie lub w jego zrównoważonej wersji.

W analizowanych orzeczeniach nie ma zbyt wielu informacji na temat procedur pobierania danych korpusowych, co nie znaczy, oczywiście, że pominięte zostały one także w opiniach biegłych. Jako że treść orzeczenia ściśle wynika z powtarzalnej struktury tekstu, w jego treści nie znajdziemy przywołanych w całości fragmentów opinii biegłych dotyczących sposobu wykorzystania korpusu, ale jedynie te odwołania, które odpowiadają punktom zaskarżenia konkretnego wyroku. Przykładowo jedno z orzeczeń zawiera natomiast dość szczegółowy opis działania, który obejmuje dwie daty korzystania ze źródeł korpusowych oraz gramatyczne i liczbowe informacje o wynikach:

Ww. aneks został oparty także na materiałach pozyskanych z korpusów języka polskiego w dniu 12 lutego 2017 r. W 52 wynikach poszukiwań słowo „(...)” występuje w połączeniu ze słowem (...) w różnych odmianach przez przypadki (tworząc pełną nazwę tego miasta – (...)) 50 razy; 1 raz w wyrażeniu „stary gród (...)” oraz 1 raz w wyrażeniu „Rynek (...)”, bez odniesienia do jakiegokolwiek miasta. W wynikach znalezionych w Narodowym Korpusie Języka Polskiego w dniu 13 lutego 2017 r. badająca przedstawiła 29 wyników. Dwanaście z nich w sposób nie budzący wątpliwości dotyczy rodzaju sądów a jeden (...). Pozostałe wyniki nie wskazują jednoznacznie czego dotyczy użyte słowo „trybunalskie” (Trybunalskie1/WSA/2019).

Ze względu na to, że NKJP nie jest korpusem monitorującym, pobieranie danych w różnych terminach przy tych samych ustawieniach nie powinno mieć wpływu na wynik. Z pewnością jednak istotna jest informacja, czy wyszukiwanie obejmowało tylko formę podstawową wyrazu czy też jego wszystkie formy fleksyjne (z dodaniem do wyrazu podwójnej gwiazdki), pojedynczy wyraz czy też jego połączenia z zachowaniem wskazanego szyku i odstępu.

Wskazany zestaw informacji o źródle i sposobie jego wykorzystania może być niezbędny dla sądu, gdyż – jak to zostało przypomniane w jednym z analizowanych orzeczeń:

sfera merytoryczna opinii biegłego kontrolowana jest przez sąd, który nie posiada wiadomości specjalnych” (Prenatal1/SO/2014).

## **Reprezentatywność korpusu a cel jego wykorzystania**

Reprezentatywność korpusu jest jedną z jego najważniejszych cech, która decyduje o wartości empirycznej materiału. Wśród różnych definicji reprezentatywności



korpusu znajdziemy, m.in. takie, które z reprezentatywnością korpusu wiążą: brak selekcjonowania danych; wielkość zasobu, który może być uznany miniaturową wersją języka, całkowicie pokrywającą jego zróżnicowanie; zrównoważenie zbioru; możliwość poprawnego szacowania oraz odpowiedzi na konkretne pytanie badawcze (Egbert, Biber, Gray, 2022, s. 31–38).

W analizowanych orzeczeniach nie pojawiały się zazwyczaj uwagi na temat reprezentatywności wykorzystanego korpusu. Jednokrotnie tylko nadawca wypowiedział się na temat zawartości korpusu, która została zakwestionowana przez stronę. W niniejszym postępowaniu na podstawie opinii biegłej, która wyjaśniła budowę korpusu, Sąd Apelacyjny nie podzielił zarzutu nieprzydatności opinii, mimo że skarżący podważał ją ze względu na zawartość w materiale korpusowym:

książek s-f, literatury kobiecej, podręcznika o tarocie, oraz artykułów z „Super Expressu” (Zabójca1/SA/2015)<sup>9</sup>.

Reprezentatywność korpusu uznaje się za właściwość całego zasobu, ale możliwe jest też wskazanie reprezentatywności zbioru, która umożliwi odpowiedź tylko na konkretne pytania badawcze: korpus będzie nadawał się do analizy wyrazów funkcyjnych, które mają bardzo dużą frekwencję, ale nie da odpowiedzi na pytania związane z innymi zagadnieniami leksykalnymi. Stąd też interesujące jest, jakie szczegółowe problemy relacjonowane w orzeczeniach były rozwiązywane dzięki odwołaniom do korpusów.

Zebrane orzeczenia sądowe powiązane są z kilkoma różnymi zagadnieniami językowymi, które sprowadzają się głównie do trzech kwestii: ustalania znaczenia wyrazu, jego nacechowania czy datowania.

Problem znaczenia wyrazu może dotyczyć występowania danego wyrazu w różnych w odmianach stylowych polszczyzny, co wiąże się z odmiennością jego znaczeń. W analizowanych orzeczeniach materiał korpusowy pomógł w potwierdzeniu różnicy między rozumieniem jednostki leksykalnej *zabójca* na gruncie języka prawa oraz w języku potocznym<sup>10</sup>. Zderzenia tych dwóch stylów są istotne nie tylko w ramach orzeczeń do pojedynczych postępowań, ale dotyczą w ogóle formułowania aktów prawnych, *którym należy nadawać znaczenie występujące w języku potocznym* (por. Choduń, 2004, s. 77), określanym też mianem powszechnego czy codziennego.

Jeszcze inny problem związany ze znaczeniem wyrazu to stwierdzenie, czy analizowany wyraz występuje w języku jako nazwa własna czy pospolita (por. *cinikciarz*, *prenatal*, *mixelko*). Źródła leksykograficzne, a zwłaszcza słowniki czy korpusy ogólne nie odnotowują w sposób systematyczny nazw własnych (Żmigrodzki, 2021), ale korpusy mogą dokumentować proces apelatywizacji tych nazw (Piwowar, 2019, s. 285), określanej także mianem deonimizacji lub nazywanej degeneracją znaku towarowego (Graf, 2021, s. 290). W tego typu sprawach istotne jest uwzględnienie, z jakiego okresu pochodzi lub jakich czasów dotyczy ilustracja przykładowa

<sup>9</sup> Por. uwagi na temat NKJP – (Przepiórkowski i in. 2012) i (Andrzejczuk 2010).

<sup>10</sup> Por. (Pawłowski 2003, s. 168).



konkretnego znaczenia, a takie informacje zawiera np. NKJP, który uwzględnia bardzo obszerne datowane cytaty, por.:

16 marca 1989 r. otworzył podwoje pierwszy kantor wymiany walut. Kończył epokę waluciarzy i cinkciarzy, choć w swej istocie był legalnym przedłużeniem tej nielegalnej w PRL profesji. (Polityka 14.03.2003, źródło – NKJP) – Cinkciarz1/NSA/2019.

Ten przykład jest potwierdzeniem ewolucji znaczenia, która wynika z obiektywnych zmian w rzeczywistości pozajęzykowej, więc jest stosunkowo łatwa do datowania. NKJP nie jest zrównoważonym korpusem diachronicznym, ale umożliwia wygenerowanie profili czasowych dla analizowanych jednostek (Przepiórkowski i in., 2012, s. 310), co może być pomocne w ogólnym datowaniu zmian znaczenia nie tylko na podstawie konkretnych ustaleń jakościowych, ale też danych frekwencyjnych. NKJP jednak, mimo że uwzględnia nazwy własne, to nie proponuje anotacji nazw produktów. W tym przypadku potrzebna jest bowiem szczegółowa analiza kontekstów, gdyż nawet odwołania do konkretnego produktu w przykładzie nie mogą potwierdzić, czy poszukiwany wyraz sytuuje się w granicach nazwy własnej czy też tylko pojawia się na etykiecie produktu i pełni funkcję nazwy opisowej, co zostało podkreślone w orzeczeniu na podstawie opinii biegłego (Prenatal1/SO/2014; Prenatal2/SA/2014). Postępowania związane ze znakami towarowymi dotyczą natomiast nie tylko statusu jednostki leksykalnej jako nazwy pospolitej lub własnej, ale także kwestii bardziej szczegółowych, np. różnic między nazwą opisową, rodzajową a fantazyjną<sup>11</sup>, np.:

Co więcej, mimo iż w korpusie języka polskiego wyraz ten w większości występuje jako nazwa własna, to jednakże w 6 na 26 wystąpień został użyty bez powiązania z konkretnym produktem czy działalnością gospodarczą, a więc w funkcji opisowej (Prenatal2/SA/2014).

Lp.	Używana w artykule nazwa orzeczenia	Cel wykorzystania korpusu
1.	Mixełko1/WSA/2012 Mixełko2/NSA/2013	Sprawdzenie, czy w zasobach korpusowych został odnotowany szukany wyraz. To miałyby potwierdzać status badanej jednostki leksykalnej jako nazwy rodzajowej, a nie własnej (tu: fantazyjnej).
3.	Prenatal1/SO/2014 Prenatal2/SA/2014	Ustalenie, czy przymiotnik funkcjonuje w korpusie jako nazwa opisowa czy własna
4.	Zabójca1/SA/2015	Sprawdzenie znaczenia rzeczownika zabójca w znaczeniu potocznym, a nie specjalistycznym

<sup>11</sup> Por. (Graf, 2021, 290–293).

6.	Zniesławienie1/SO/2018	Sprawdzenie nacechowania wyrazów, które zostały uznane za obraźliwe
7.	Trybunalskie1/ WSA/2019	Sprawdzenie, czy w świadomości użytkowników języka analizowany przymiotnik trybunalskie powiązany jest z nazwą miasta czy z rzeczownikiem pospolitym trybunał.
8.	Cinkciarz1/NSA/2019	Występowanie wyrazu cinkciarz w języku potocznym i ustalenie jego współczesnego znaczenia

Tab. 2. Cel wykorzystania korpusów – na podstawie analizowanego materiału

## Wnioski

Na koniec warto zauważyć, że korpusy mogą być wykorzystywane w orzecznictwie na dwa sposoby. Pierwszy to traktowanie korpusu jako ogromnego słownika, który będąc dokumentem elektronicznym, zawiera wszystkie informacje niemieszczące się w papierowej publikacji, np. ogromną liczbę bardzo obszernych przykładów opatrzonych metadanymi<sup>12</sup>.

Ta metoda ma wymiar analizy jakościowej, dostępnej zarówno dla biegłych, jak i dla prawników. Natomiast drugi sposób argumentacji korpusowej to traktowanie tego zasobu jako źródła danych ilościowych, które powinny być przeanalizowane przez osobę dysponującą tzw. wiadomościami specjalnymi<sup>13</sup>. Tylko językoznawca może przeprowadzić tego typu badanie zgodnie z najnowszą wiedzą na temat metodologii korpusowej, świadomości jej zalet i ograniczeń. W analizowanych orzeczeniach pojawiły się także odniesienia do danych liczbowych, które nie mogą być jednak uznane za przykład zastosowania metod ilościowych, jeśli nie znamy np. zakresu przeprowadzanej analizy, która stałaby się podstawą porównania dla uzyskanych wyników (Piotrowski, Grabowski, 2013).

Tego typu wyjaśnienia nie pojawiały się jednak w orzeczeniach, gdyż źródła korpusowe były wykorzystywane przede wszystkim z perspektywy jakościowej, a nie ilościowej.

Analiza korpusowa może przynieść obiektywne, rzetelne i wiarygodne dane, które staną się wartościowym materiałem w podejmowaniu decyzji w ramach orzecznictwa. Zadaniem biegłego jest takie przedstawienie procedury badania, jej objaśnienie i wizualizacja, aby mogły stanowić punkt odniesienia dla sędziego. Warto zauważyć,

<sup>12</sup> Por.: Although corpora are not literally dictionaries, they have an important functional similarity in that they contain not merely a few illustrations of word usage but potentially thousands of examples. (Stempel 2021, s. 396).

<sup>13</sup> Por. uwagi na temat wykorzystywania słowników przez lingwistów i nielingwistów (Kubicka, Grec, Leszczak, Mitéw, 2015, s. 26–28).

ze analiza korpusowa nie jest z pewnością procedurą trudniejszą do zrozumienia niż wykorzystywane w orzecznictwie dane z badań genetycznych czy fizyko-chemicznych.

Aby ten drugi sposób wykorzystania źródeł korpusowych mógł na większą skalę funkcjonować w polskim sądownictwie, konieczne jest funkcjonowanie korpusów w świadomości prawników. Stanie się tak wtedy, gdy w polskim orzecznictwie zadomowi się pierwszy, jakościowy sposób wykorzystywania korpusów, a dzięki temu korpusy zaczną budzić wśród prawników zaufanie jako dostępne źródło obiektywnych danych językowych.

### **Bibliografia**

- Andrzejczuk, A., (2010). Narodowy Korpus Języka Polskiego – teoria i praktyka. Fakty, mity, potrzeby. *Comparative Legilinguistics*, 3, 133–142.
- Bernstein A. (2021). Legal Corpus Linguistics and the Half-Empirical Attitude. *Cornell Law Review*, 106, 1397–1456.
- Bielska-Brodziak, A., Tobor, Z., Żmigrodzki, P. (2008). Co każdy prawnik o słownikach wiedzieć powinien. *Przegląd Sądowy*, 7–8, 79–95.
- Choduń, A., (2004). Język prawny a język potoczny. W: A. Malinowska (red.) *Język – prawo społeczeństwo*. Opole: Wydawnictwo Uniwersytetu Opolskiego.
- Choduń, A., Zieliński, M. (2008). W sprawie słowników w procesie interpretacji tekstów prawnych raz jeszcze. *Państwo i Prawo*, 7, 96–100.
- Czelakowska, A., (2016). Nieufnym okiem prawnika, czyli najnowszy (elektroniczny) słownik języka polskiego Polskiej Akademii Nauk z punktu widzenia wykładowi językowej. *Kwartalnik Sądowy Apelacji Gdańskiej*, 2, 7–27.
- Czelakowska, A., Kubicka, E., Klubińska, M. (2016). Słowniki przed sądem. Wykorzystanie dzieł leksykograficznych w orzeczeniach sądowych. *LingVaria*, 21, 45–64.
- Egbert, J., Biber, D., Gray, B. (2022). *Designing and Evaluating Language Corpora. A practical Framework for Corpus Representativeness*. Cambridge: Cambridge University Press.
- Graf, M. (2021). Onomastyka i prawo. Miejsca wspólne. *Biuletyn Polskiego Towarzystwa Językoznawczego*, 77, 279–295.
- Gizbert-Studnicki, T. (2007). *Postępowanie sądowe jako złożony dyskurs – zagadnienia pragmatyki i języka sali sądowej*. www.rjp.pan.pl, Kraków.
- Goźdz-Roszowski, S. (2021). Corpus linguistics in legal discourse. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique Juridique*, 34 (5), 1515–1540.
- Gries, S.Th. (2021). Corpus linguistics and the law: extending the field from a statistical perspective. *Brooklyn Law Review*, 86(2), 321–356.
- Hebal-Jeziarska, M. (2013). Podstawowe zasady korzystania z korpusów przy badaniu języka. W: W. Chlebda (red.) *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*. (s. 17–30). Opole: Wydawnictwo Uniwersytetu Opolskiego.
- ISJP – Bańko, M. (red.). 2000. *Inny słownik języka polskiego*. Warszawa: PWN.

- Kubicka, E., Grec, K., Leszczak, W., Mitew, M. (2015). Wykorzystanie słowników w interpretacji prawniczej – implikacje praktyczne, *Kwartalnik Sądowy Apelacji Gdańskiej*, 1, 25–50.
- Lizisowa, M.T. (2018). *Komunikacyjna teoria języka prawnego*. Poznań: Wydawnictwo Naukowe Contact.
- Nowak-Korcz, P. (2020). Makrostruktura tekstów orzeczeń sądowych na przykładzie polskich i francuskich wyroków. W: S. Goźdź-Roszkowski, A. Buras-Marciniak (red.) *Języki specjalistyczne w komunikacji interkulturowej*. (s. 33–47). Łódź: wydawnictwo Uniwersytetu Łódzkiego.
- Pawłowski, A. (2003). Uwagi na temat korpusu języka polskiego (reprezentatywność, aktualność, nazwa). W: S. Gajda (red.), *Językoznawstwo w Polsce. Stan i perspektywy* (s. 162–180). Opole: Polska Akademia Nauk – Komitet Językoznawstwa, Uniwersytet Opolski – Instytut Filologii Polskiej.
- Piwowar, J. (2019). Zastosowanie korpusów językowych w analizach identyczności znaków słownych. W: M. Zaśko-Zielińska, K. Kredens (red.), *Lingwistyka kryminalistyczna. Teoria i praktyka* (s. 275–288). Wrocław: Uniwersytet Wrocławski, Quaestio.
- Piotrowski, T., Grabowski Ł. (2013). Interpretacja danych frekwencyjnych z korpusów językowych: opis pewnych problemów (na kilku przykładach z życia wziętych). W: W. Chlebda (red.), *Na tropach korpusów: w poszukiwaniu optymalnych zbiorów tekstów* (s. 59–71). Opole: Wydawnictwo Uniwersytetu Opolskiego.
- Podoliński, D. (2018). Kategoria słuszności i jej leksykalne wykładniki na podstawie uzasadnień wyroków białostockich sądów apelacyjnych, *Białostockie Archiwum Językowe*, 18, 181–195. /doi: 10.15290/baj.2018.18.12
- Przepiórkowski, A., Bańko, M., Górski, R. L., Lewandowska-Tomaszczyk, B. (red.) (2012). *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Rzucidło-Grochowska, I. (2017). Strategie i techniki formułowania uzasadnień orzeczeń sądowych. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, 2, 59–72.
- Stempel, J. W. (2021). *Adding Context and Constraint to Corpus Linguistics*. *Brooklyn Law Review*, 86(2). 389–434.
- Suska, M. (2019). Internetowe słowniki polszczyzny w procesie wykładni prawa. W: S. Tkacz, Z. Tobor (red.), *Prawo a nowe technologie*. (s. 52–60). Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Tobor, Z., Zeifert, M. (2021). Linguam novit curia? Analiza korpusowa w prawie karnym – przyczynek do dyskusji. *Forum Polityki Kryminalnej*, 1, / doi.org/10.31261/FPK.2021.01.04
- Tobor, Z., Zeifert, M. (2020). Korpusy językowe jako narzędzie interpretacji prawa: amerykańska teoria i praktyka. *Archiwum Filozofii Prawa i Filozofii Społecznej*, 4, 80–90.
- Zaśko-Zielińska, M. (2013). Listy pożegnalne: w poszukiwaniu lingwistycznych wyznaczników autentyczności tekstu. Wrocław: Wydawnictwo Quaestio.
- Zaśko-Zielińska, M. (2021). Interdyscyplinarne pomosty między językoznawstwem a prawem w lingwistyce kryminalistycznej. W: M. Guławska, G. Pawłowski (red.), *Lingwistyka à la inter: status i perspektywy badań interdyscyplinarnych*. (s. 111–123). Warszawa: Uniwersytet Warszawski.

- Żmigrodzki, P., (2021). Nazwy własne i wyrazy od nich pochodne w *Wielkim słowniku języka polskiego PAN*. Plany, stan obecny, perspektywy. *Onomastica*, 65, s. 41–53.
- Żurowski, S. (2014). Leksykografia w pracy prawnika. W: E. Kubicka, D. Kala (red.), *Kultura języka polskiego w praktyce prawniczej* (s. 55–75). Bydgoszcz: Zrzeszenie Prawników Polskich. Oddział: Qartis.

### **Źródła internetowe**

- NKJP – Narodowy Korpus Języka Polskiego: <http://www.nkjp.pl/>
- Legalis – <https://www.beck.pl/system-informacji-prawnej-legalis/>
- [https://for.org.pl/upload/Nowy\\_Wymiar\\_Sprawiedliwosci/Raport\\_Publiczna\\_dostepnosc\\_orzeczen\\_sadowych.pdf](https://for.org.pl/upload/Nowy_Wymiar_Sprawiedliwosci/Raport_Publiczna_dostepnosc_orzeczen_sadowych.pdf)

### **Słowa kluczowe**

orzeczenie sądowe, korpus, reprezentatywność

### **Abstract**

#### **Corpora and digital tools in linguistic analyses of legal problems. References to corpora in the judicial decision of Polish courts**

The aim of the article is to analyze court decisions in terms of the ways they make use of the methodology of corpus linguistics. With this aim in mind, the following text properties have been examined: the way the information about the source of the corpus is provided in the rulings of Polish courts, what problems reported in the rulings were solved with the use of corpora, and what comments on the use of corpora are found in the examined documents. The analysis of court decisions shows that they often lack information necessary to identify corpora. There is also no mention of the assumed rules for data collection. The corpora are used primarily in the qualitative perspective but the correct application of quantitative methods may allow generating more objectified data helpful in assessing the linguistic material.

### **Keywords**

court decision, corpus, representativeness