

KAROLINA LISZYK

MOŻLIWOŚCI I OGRANICZENIA NARZĘDZI CYFROWYCH  
W BADANIACH  
NAD SŁOWNICTWEM NACECHOWANYM AKSJOLOGICZNIE  
(NA PRZYKŁADZIE CZASOWNIKA *GORSZYĆ*)

Zwrot cyfrowy<sup>1</sup>, który nastąpił w polskiej humanistyce przed kilkunastu laty, otworzył przed badaczami nowe możliwości. Dla lingwistów wzbogacenie aparatu metodologicznego o zaplecze technologiczne oznacza zarówno dostęp do dużych baz danych językowych, jak i możliwość korzystania z wielu narzędzi do szybkiej i różnorodnej analizy materiału badawczego. Niemal od zarania rodzimej lingwistyki cyfrowej obszarem najmniej podatnym na rewolucję informatyczną była jednak semantyka, na co zwracali uwagę zarówno językoznawcy, jak i inżynierowie języka, np.:

Jak pokazuje doświadczenie, semantyka pozostaje dla inżynierii języka obszarem trudnym, kosztownym w badaniach i niepewnym pod względem weryfikacji, o którym mówi się eufemistycznie, że „stanowi wyzwanie”, co w przekładzie na otwarty tekst oznacza, iż jest barierą nie do pokonania. (Pawłowski i in., 2010, s. 56)

Semantyka pozostawała przez lata obszarem niedostępnym badaniom ilościowym, ponieważ wymagają one danych obserwowalnych empirycznie i przeliczalnych, a pojęcia „sensu” czy „znaczenia”, niezależnie od teorii źródłowej określającej ich zakres, były przez długi czas niemierzalne. (Pawłowski, 2020, s. 244)

---

Dr KAROLINA LISZYK – Uniwersytet Śląski w Katowicach, Wydział Humanistyczny, Instytut Językoznawstwa; adres do korespondencji: ul. Uniwersytecka 4, 40-007 Katowice; e-mail: karolina.liszyk@us.edu.pl; ORCID: <https://orcid.org/0000-0002-8741-1595>.

<sup>1</sup> Wyrażenie to zostało użyte w nazwie konferencji, która jako pierwsza w Polsce dotyczyła problematyki humanistyki cyfrowej. Symposium „Zwrot cyfrowy w humanistyce” odbyło się 25-26 października 2012 roku w Centrum Języka i Kultury Polskiej w Lublinie, a jej pokłosiem jest tom *Zwrot cyfrowy w humanistyce. Internet. Nowe media. Kultura 2.0* pod redakcją Andrzeja Radomskiego i Andrzeja Bomby. O samym terminie *humanistyka cyfrowa* i sposobach jego konceptualizowania zob. także Duda i in., 2019.

---

Artykuły są objęte licencją Creative Commons Uznanie autorstwa – Użycie niekomercyjne – Bez utworów zależnych 4.0 Międzynarodowe (CC BY-NC-ND 4.0)

---

Obecnie językoznawcy zajmujący się badaniem znaczenia i użycia słów mogą pracować na materiale wyekscerpowanym z różnych korpusów językowych, a także korzystać z oferowanych przez nich funkcji, np. kolokatorów, profili diachronicznych czy wykazu źródeł sporządzanego na podstawie metadanych. Poszerzenie analiz semantycznych możliwe jest także dzięki programom do automatycznego wydobywania z analizowanego materiału językowego list frekwencyjnych słownictwa, słów kluczowych czy w końcu określania tematyki dużych tekstów przy zastosowaniu narzędzi do modelowania tematycznego<sup>2</sup>.

Za cel obieram sobie zestawienie wyników analiz przeprowadzanych przy użyciu wymienionych rozwiązań z zakresu lingwistyki cyfrowej z wnioskami wysnuwanymi w toku subiektywnej interpretacji. Przedmiotem owych dociekań czynię próbę wskazania kontekstów, w których używany jest współcześnie czasownik *gorszyć* – określenia przede wszystkim jego cech semantycznych, poprzez wskazanie konkretnych postaw lub działań opisywanych w ten sposób (innymi słowy: kto lub co, a także czym lub jak nas gorszy).

Wybór badanego słowa nie jest przypadkowy – notowane jest ono w polszczyźnie od XVI wieku (por. Boryś, 2005, s. 172) i konsekwentnie definiowane w słownikach ogólnych polszczyzny za pomocą bliskoznaczników, m.in. *demoralizować*, *dawać zły przykład*, *wywierać zły wpływ*, *wywoływać sprzeciw/oburzenie/niezadowolenie*<sup>3</sup>. Objaśnianie znaczenia słownictwa nacechowanego aksjologicznie nie jest łatwe, wymaga bowiem odwołań do norm społecznych, moralnych, obyczajowych, religijnych czy prawnych, które – zwłaszcza w odniesieniu do trzech pierwszych typów – mają najczęściej charakter umowny, niepisany, są warunkowane przynależnością do określonej grupy społecznej, etnicznej czy terytorialnej i mogą podlegać transformacjom wynikającym z przeobrażeń kulturowych lub szeroko pojętych doświadczeń dziejowych. To, co w definicjach oznaczane jest wyrażeniami: *zły przykład*, *demoralizujący wpływ* czy *normy obyczajowe i moralne*, pozostaje dość enigmatyczne – chociaż można bowiem mówić o normach uniwersalnych (zakłada się, że te, które zaspokajają elementarne

---

<sup>2</sup> Wykaz ten nie wyczerpuje listy dostępnych narzędzi do automatycznej analizy tekstu. Wybrałam te, które w mojej ocenie są najczęściej wykorzystywane przez językoznawców.

<sup>3</sup> Por. definicje z wybranych słowników – SPXVI: ‘W odniesieniu do moralności: dawać zły przykład, demoralizować, wywierać ujemny wpływ’; SL (podobnie SWil): *moral*. ‘do złego pobudzać przykładem, mową etc.’; SWar: ‘pokazywać komu drogę do złego, dawać zgorszenie, dawać zły przykład, psuć demoralizować’; SJPD: ‘wywierać zły, demoralizujący wpływ; powodować zepsucie, zgorszenie; wywoływać oburzenie’; ISJP: ‘Jeśli coś gorszy jakąś osobę albo jeśli ona gorszy się czymś, to narusza to jej poczucie przyzwoitości, ponieważ jest niemoralne lub złe.’; WSJP PAN: 1. ‘wywoływać sprzeciw lub niezadowolenie przez niestosowanie się do powszechnie przyjętych norm obyczajowych i moralnych’ 2. ‘wywierać demoralizujący wpływ’.

potrzeby ogółu, muszą być wspólne – zob. Ossowska, 1957), to jednak nie sposób sporządzić ich kompletnego zestawienia, obowiązującego zawsze i wszędzie.

Dane leksykograficzne stanowią jednak dla semantyków źródło nie do przecenienia – w erze przed rewolucji cyfrowej (choć także i dziś) ich rzetelna analiza była jednym z wiodących sposobów dociekania znaczenia. W odniesieniu do czasownika *gorszyć* próbę taką podjęła Renata Grzegorzczkova, która wskazała jego rozwój semantyczny od staropolszczyzny do współczesności. Badania te doprowadziły uczoną do wniosku, że u progu XXI wieku czasownik *gorszyć* używany był jako synonim czasownika *demoralizować* – ‘powodować w kimś powstanie zła moralnego’ – oraz jako wtórne *causativum* do *gorszyć się* – ‘przeżywać, wyrażać oburzenie moralne’ (Grzegorzczkova, 2000, ss. 216-217). To bardzo cenne konkluzje, jednak bez odniesienia do dużego zbioru autentycznych zdań niemożliwe pozostaje wskazanie konkretnych zachowań ilustrujących, co znaczy *gorszyć*.

Jadwiga Puzynina postulowała, by lingwiści pytali, jak ludzie mówiący określonym językiem w pewnej epoce rozumieją słowa nazywające wartości (Puzynina, 1982, s. 25). W realizacji tego zadania niezwykle pomocne są współcześnie korpusy językowe. Dostęp do dużych baz danych językowych z jednej strony ułatwia lingwistom pracę, umożliwiając szybkie pozyskanie zbioru autentycznych wypowiedzi zawierających wyszukiwane słowoformy, z drugiej jednak – ewokuje nowe problemy:

W przypadku wypowiedzi językowej zapisanej w postaci fragmentu tekstu w korpusie, dostępna wiedza o jej kontekście użycia jest bardzo uboga, np. często znamy autora, ale sytuacja, adresat, czas, miejsce itd. są zwykle nieokreślone lub określone bardzo nieprecyzyjnie. Rodzą się pytania, jak bardzo ogranicza to możliwość poznania znaczenia wyrażen językowych poprzez analizę zawężoną do zapisów wypowiedzi językowych zawartych w korpusach tekstu, oraz w jakich zastosowaniach i obszarach tak pozyskana wiedza o znaczeniu może być pomocna? (Piasecki, 2010, s. 143)

Pozyskanie materiału korpusowego, mimo potencjalnych problemów związanych z jego interpretacją, wydaje się krokiem koniecznym, by na dalszym etapie prac korzystać z narzędzi do automatycznej analizy danych językowych. W moich badaniach oparłam się na zasobach korpusu referencyjnego MoncoPL (Pęzik, 2020), kierując się przede wszystkim aktualnością ekscerpowanych wypowiedzi (uwzględniane przeze mnie konkordancje pochodzą z okresu 1 stycznia 2008–30 kwietnia 2023).

Próba wyszukania zbioru zdań obejmujących wszystkie warianty fleksyjne czasownika *gorszyć* przyniosła wynik jedynie częściowo satysfakcjonujący. Choć bowiem wyodrębniony zbiór okazał się dość zasobny (15 306 konkordancji),

to ostatecznie dalszym badaniom poddanych zostało jedynie 769 wypowiedzi. Przyczynami, dla których konieczne było zawężenie pozyskanego materiału, są wieloznaczność wyszukanych form tekstowych i ich niewłaściwa interpretacja dokonana przez analizator morfologiczny – w wyodrębnionym zbiorze znakomita większość konkordancji zawierających słowoformy *gorszy* lub *gorszą* to zdania, w których stanowią one formy stopnia wyższego przymiotnika *zły*. Proces ręcznej dezambiguacji, przeprowadzanej w pliku pobranym ze strony korpusu, posłużył wyeliminowaniu realizacji wyrazowych innych leksemów, także przykładów z czasownikiem *gorszyć się* oraz z imiesłowami przymiotnikowymi (służą wskazywaniu cech określanych nimi rzeczowników i wymagają innego opisu semantycznego).

Konieczność selekcji wyekscerpowanego z korpusu materiału pociąga za sobą pewne konsekwencje – niemożliwe jest w takim wypadku skorzystanie z dodatkowych funkcji oferowanych przez twórców korpusu, pozwalających uzyskać wykazy najczęstszych kolokacji, typów źródeł, w których odnotowano wyszukiwaną jednostkę, oraz informacji o cyklach frekwencyjnych. Oczywiście, opisywana sytuacja nie jest typowa – wybór innego słowa do badań nie skutkowałby takimi ograniczeniami.

W rozwiązaniu problemu z automatycznym pozyskaniem kolokacji pomocny okazuje się Korpusomat (korpusomat.pl) – narzędzie do samodzielnego tworzenia korpusów, wzbogacone o liczne możliwości ich przeszukiwania (Kieraś, Kobyliński, 2021). Materiał wprowadzony przeze mnie do Korpusomatu składał się z 769 zdań z czasownikiem *gorszyć* wyodrębnionych z zasobów MoncoPL, pozbawionych metadanych i wgranych w postaci zbiorczego pliku (utworzony w ten sposób korpus jest niewielki – liczy 18 589 segmentów, jednak wystarczający na potrzeby realizacji założonego celu). Aby pozyskać wykaz kolokacji, korzystałam zarówno z funkcji „profile słów” (pozwalającej wygenerować częste liczby połączeń wyszukiwanego wyrazu z innymi), jak i statystyk generowanych podczas samodzielnego odpytywania korpusu. Założyłam, że nie wszystkie typy konkurencji wyrazowych są tak samo istotne w dociekaniach dotyczących znaczenia i użycia wyrazu. Przekonuje mnie w tym zakresie opinia Renaty Przybylskiej, która opisując etap gromadzenia kolokacji przydatnych do analizy semantycznej, zwróciła uwagę, że należy „wyłonić połączenia wyrazowe, które są: 1. najczęstsze, 2. poznawczo istotne dla semantyki badanego wyrazu hasłowego” (Przybylska, 2020, s. 46). Uważam, że w odniesieniu do czasownika *gorszyć* najistotniejsza jest jego obustronna łączliwość z rzeczownikami (nazywanymi podmiot, obiekt i środek czynności), a także z czasownikami, z którymi badany wyraz wchodzi w relację koniunkcji.

Z pozyskanych kolokacji wynika, że najczęściej (co najmniej pięciokrotnie) wskazywanymi obiektami akcji określanej czasownikiem *gorszyć* są kolejno: *dziecko, człowiek, maluczki, inny, wierny, społeczeństwo, mieszkaniec, młodzież, widz, przechodzień, opinia publiczna, publiczność, poddany, czytelnik, uczeń* oraz *katolik*. Wyniki te da się podzielić na kilka grup: osoby nieletnie (to najliczniejsza klasa – zdania, w których mowa o gorszeniu dzieci i młodzieży można więc uznać za wyraz troski i wolę ochrony tych, których system moralny dopiero się kształtuje), osoby będące świadkiem jakiegoś wydarzenia (intencjonalnie – jak *widz, publiczność, czytelnik* – lub przypadkowo, wskutek przebywania w określonym miejscu – *mieszkaniec, przechodzień*), osoby wyróżnione ze względu na ich wyznanie oraz reprezentantów bliżej niedookreślonych zbiorowości. Świadomie wskazują przy tym, że mowa o dużej liczbie osób – dla wszystkich wymienionych dopełnień czasownika *gorszyć* charakterystyczna jest bowiem ich liczba mnoga. Jej użycie jest sygnałem tego, że działania lub postawy, o których mowa w poszczególnych wypowiedziach, nie są skierowane na jednostki – przeciwnie, zagrażają (lub mogą zagrażać) poczuciu moralności wielu osób<sup>4</sup>.

Trudniejszą do wyodrębnienia grupą są nazwy wykonawców czynności, czyli tych, którzy gorszą, najczęściej bowiem w analizowanych zdaniach pozycję składniową agensa zajmuje nazwa samego działania lub postawy, nie zaś jej subiektu (nierzadko jednak możliwego do wyinterpretowania po lekturze szerszego kontekstu). Z materiału korpusowego wynika, że biorąc pod uwagę jedynie kookurencję wyrazów do osób, które gorszą, zaliczyć można najczęściej *księży* oraz przed dwoma tysiącami lat *Jezusa*. Jeśli natomiast uwzględnić rzeczowniki pełniące funkcję podmiotu w zdaniach z czasownikiem *gorszyć*, to tym, co gorszy, są: *zachowanie, grzech, słowo, dowcip* i *zdjęcie*. Rzeczowniki te, wyjęte z kontekstu, niewiele wnoszą – nazywają jedynie sposób realizacji gorszenia. Jego wykładników można także poszukiwać wśród narzędnikowych form rzeczowników łączących się z czasownikiem *gorszyć*, choć i w tej grupie występują ponownie ogólnikowe kolokanty *zachowanie* i *grzech*.

Istotne dla rozważań semantycznych są także kolokacje, w których formy wyrazu *gorszyć* wchodzą w relację koniunkcji z innymi czasownikami. Jeśli założyć, że połączenia te nie są redundantne, to zakresy tworzących je słów nie pokrywają się całkowicie. Współwystępowanie czasowników stanowi wówczas próbę wzajemnego dookreślenia i uzupełnienia ich znaczeń. W moim materiale

---

<sup>4</sup> Objęcie akcją gorszenia dużych zbiorowości szczególnie dobrze widać w nieanalizowanych tu kolokacjach z rzeczownikiem *zgorszenie*, np. *siac, szerzyć zgorszenie*. Jedną z wiązek semantycznych tych czasowników jest wszak rozpowszechnianie, rozprzestrzenianie tego, co wyrażają rzeczowniki łączące się z nimi w funkcji dopełnienia.

badawczym najczęstsze są szeregi, w których czasownik *gorszyć* zestawiany jest z wyrazami: *szokować, dziwić, obrażać, fascynować, oburzać, ranić, zakłócać spokój/porządek, wywoływać skandal, bulwersować, deprawować* oraz *zaskakiwać*. Można więc przypuszczać, że działania lub postawy związane z gorszeniem kogoś mają charakter niespodziewany i wzbudzają gwałtowne, negatywne emocje<sup>5</sup>.

Z oglądu automatycznie wydobywanych kolokacji, pozbawionych kontekstu i niepopartych interpretacją, wynikają pewne przesłanki dotyczące znaczenia i zakresu użycia czasownika *gorszyć*, jednak tylko dane odnośnie do obiektu aktu gorszenia są precyzyjne, reszta informacji pozostaje dość ogólna. Z pomocą przychodzą kolejne funkcje oferowane przez narzędzia cyfrowe.

Nieco więcej pozwala dostrzec analiza generowanej przez Korpusomat listy frekwencyjnej całego słownictwa stanowiącego zawartość korpusu (możliwa także do pozyskania np. przy zastosowaniu programu AntConc), a więc obejmująca także te wyrazy, które nie wchodzą w bezpośrednie zależności syntaktyczne z czasownikiem *gorszyć*. Wśród najczęściej odnotowanych można wskazać dużą grupę słownictwa odnoszącego się do tematyki religijnej<sup>6</sup>: *Kościół, ksiądz, Jezus, grzech, wierny, Bóg, wiara, biskup, Pan, katolik, papież, krzyż, chrześcijanin, Ewangelia, sumienie, grzesznik, faryzeusz, msza, wspólnota, poganin, Chrystus, proboszcz* oraz przymiotniki: *święty, boży, chrześcijański, religijny, kościelny*. Drugą grupę stanowi słownictwo dotyczące szeroko pojętej kultury i sztuki, choć wieloznaczność niektórych słów (np. *dzieło, obraz*) skłania do nieprzyjmowania takiego wniosku za pewny bez weryfikacji materiału. Kwalifikuję tu rzeczowniki: *dzieło, książka, zdjęcie, obraz, teatr, scena, muzeum, kultura, czytelnik, widz, film, publiczność, rzeźba*. Trzecia grupa to z kolei słowa odnoszące się do fizyczności człowieka: *seksualny, ciało, pierś, nagi, seks, nagość*. Wygenerowana lista frekwencyjna pozwala więc dostrzec pewne pola leksykalne, choć słownictwo w nich zgromadzone w różny sposób dotyczy aktu gorszenia – wskazuje na wykonawcę czynności, obiekt, środek przekazu lub okoliczności zdarzenia.

<sup>5</sup> Dopelnieniem tej kwestii jest ogląd materiału porównawczego, wyekscerpowanego z korpusu równoległego polsko-angielskich i angielsko-polskich tłumaczeń – Paralela (Pęzik, 2016). Tylko 15 zdań z czasownikiem *gorszyć* ma tu swoje angielskie odpowiedniki (pomijam wypowiedzenia z imiesłowami przymiotnikowymi), jednak nawet ten niewielki zbiór przynosi ciekawe obserwacje – do ekwiwalentów czasownika *gorszyć* zalicza się słowa: *to shock* ('szokować'), *to scandalize* ('wywoływać skandal'), *to confuse* ('wprawiać w zakłopotanie'), *to demoralise* ('demoralizować'), *to distress* ('martwić'), *to destroy* ('niszczyć') oraz *to freak out* ('przerażać'). Zestaw ten pokazuje, że polski czasownik *gorszyć* łączy się z zaskoczeniem i negatywną oceną czyjegoś zachowania lub postawy.

<sup>6</sup> Wymieniam słowa w kolejności ich występowania na liście frekwencyjnej, zaczynając od najliczniej reprezentowanych, a kończąc na wyrazach, które odnotowane zostały co najmniej pięciokrotnie.

Potwierdzeniem istotności słownictwa wskazanego w pierwszej grupie jest z kolei wykaz słów kluczowych, możliwy do wygenerowania za pomocą takich narzędzi, jak Respa, AntConc czy Korpusomat. W tym ostatnim słowa kluczowe są wydobywane z korpusu poprzez zestawienie jego listy frekwencyjnej z listą frekwencyjną korpusu referencyjnego (milionowej próbki z Narodowego Korpusu Języka Polskiego). W wykazie 50 słów zakwalifikowanych przez Korpusomat jako nacechowane największą istotnością aż 14 dotyczy tematyki religijnej (wymienione wcześniej przy omawianiu list frekwencyjnych). Uwagę zwracają także czasowniki, wskazywane już przy opisie kolokacji (*szokować, dziwić, obrażać, fascynować*), rzeczowniki nazywające obiekt gorszenia (*dziecko, młodzież, przechodzień, społeczeństwo*), a także wyrazy bezpośrednio wartościujące (*zły, zło*), dowodzące negatywnego nacechowania czasownika *gorszyć*. Jak zauważył Witold Marciszewski (1977, ss. 212–213): „w powszechnym przekonaniu (a przynajmniej odczuciu) słowa kluczowe są nam potrzebne do tego, żeby odpowiedzieć na pytanie, o czym jest mowa w danym tekście? Innymi słowy: jaki temat, czy jakie tematy są w tekście omawiane?”. Jeśli przyjąć takie założenie, to zdania z czasownikiem *gorszyć* dotyczą trzech zagadnień: zachowań lub postaw dyskutowanych w środowisku katolickim, zagrożeń dla nieletnich oraz czegoś negatywnego, co szokuje swoją odmiernością.

W odniesieniu do sygnalizowanych wniosków należy jednak mieć na uwadze, że analiza słów kluczowych powinna dotyczyć tylko tekstów koherentnych, spójnych tematycznie, nie zaś korpusów stanowiących zbiór różnorodnych (by nie rzec: przypadkowych) zdań. Podobna zasada dotyczy tekstów, które poddawane są modelowaniu tematycznemu<sup>7</sup>. Z tego powodu analiza mojego materiału badawczego za pomocą narzędzia Topic (<https://ws.clarin-pl.eu/topic>) zgodnie z przewidywaniami nie przynosi satysfakcjonujących wyników – wygenerowane struktury tematyczne nie pozwalają na jednoznaczne wyodrębnienie tematów poruszanych w zdaniach składających się na korpus.

Ciekawe informacje przynosi natomiast analiza źródeł, z których zaczerpnięte zostały przykłady składające się na materiał badawczy (co w odniesieniu do danych z MoncoPL oznacza wykaz stron internetowych)<sup>8</sup>. Biorąc pod uwagę liczbę konkordancji pozyskanych z poszczególnych portali lub serwisów, blisko 25% stanowią strony będące nośnikami treści katolickich, np. [denon.pl](http://denon.pl), [wiara.pl](http://wiara.pl), [radiomaryja.pl](http://radiomaryja.pl), [katolik.pl](http://katolik.pl), [gosc.pl](http://gosc.pl) i inne. Dla porównania: stosunek zdań, które

<sup>7</sup> Jak zaznacza Maciej Eder (2016, s. 37), jej „celem miało być wyszukiwanie treści (w tym: sporządzanie automatycznych streszczeń) w kolekcjach dokumentów, porządkowanie dokumentów pod względem ich podobieństwa tematycznego, np. w bibliotekach cyfrowych czy w centrach informacji naukowej”.

<sup>8</sup> Analizę tę przeprowadzam ręcznie, co przy objętości mojego korpusu nie jest problematyczne.

30 kwietnia 2023 roku wchodziły w skład zasobów MoncoPL i pochodzą ze stron o profilu katolickim, reprezentowanych w moim materiale, do łącznej liczby zdań tworzących korpus MoncoPL to 3%. Wypowiedzenia z czasownikiem *gorszyć* są więc w istotnym stopniu reprezentowane w dyskursie katolickim, czego przesłankami były opisane wcześniej dane pozyskane dzięki automatycznej analizie kolokacji, list frekwencyjnych oraz słów kluczowych (być może przyczyn tego należy upatrywać w nawiązaniach do leksyki ewangelicznej). Pozostałe zdania z czasownikiem *gorszyć* zaczerpnięte zostały głównie z serwisów informacyjnych i w niewielkim stopniu (4%) ze stron o charakterze rozrywkowym (np. kozaczek.pl, joemonster.pl, wykop.pl, plotek.pl, pomponik.pl), gdzie nierzadko opisywane są skandalizujące zachowania znanych osób.

Korzyści nie przynosi natomiast analiza metadanych pozwalających określić częstość występowania analizowanego czasownika w poszczególnych latach – wyniki te są porównywalne, brak tu istotnych anomalii. Taki stan rzeczy nie zaskakuje – poddawany oglądowi materiał obejmuje nieco ponad 15 lat, a to zbyt krótki okres, by mówić o tendencjach w zakresie użycia badanego słowa. Ponadto czasownik ten nie odnosi się do zjawisk, które w ostatnich latach można by uznać za wyjątkowe lub nowe.

Na podstawie danych możliwych do pozyskania dzięki automatycznej analizie materiału badawczego można więc wysnuć kilka wniosków:

1. Czasownik *gorszyć* używany jest w zdaniach związanych tematycznie z katolicyzmem: dowodzi tego wysoki odsetek materiału pochodzącego z portali publikujących wiadomości dotyczące Kościoła katolickiego; z kolokacji wynika, że księża wskazywani są jako gorszyciele, katolicy/wierni jako osoby gorszone, a grzech może być środkiem gorszenia; lista frekwencyjna wyrazów oraz zestawienie słów kluczowych potwierdzają wysoką częstotliwość oraz istotność określeń z pola tematycznego dotyczącego katolicyzmu (choć nie pozwalają na wysnuenie konkretnych wniosków, np. czy częściej mowa o katolikach jako bezpośrednich „ofiarach” gorszenia czy też ich sprawcach; które grzechy są szczególnie gorszące itd.).

2. Według autorów (zapewne dorosłych) przeanalizowanych wypowiedzi szczególnie podatną na gorszenie grupą osób są nieletni – wskazują na to kolokacje oraz wykaz słów kluczowych.

3. Akcja oznaczana czasownikiem *gorszyć* nosi znamiona aberracji – stanowi naruszenie porządku obowiązującego w określonym miejscu lub środowisku (wskazują na to kolokacje szeregowe, dane z korpusu Paralela oraz słowa kluczowe).



4. *Gorszyć* to robić coś złego – wskazują na to słowa kluczowe, ale też pośrednio inne dane (np. gorszy grzech, który w doktrynach religijnych jest wykroczeniem, a zatem oceniany jest negatywnie).

5. Trudno wskazać konkretne przejawy gorszenia, jednak na podstawie listy frekwencyjnej wyrazów oraz kolokacji można sądzić, że są to np. nagość, przełamująca konwencje twórczość (np. filmy, książki, zdjęcia) oraz – ogólniej – słowa, zachowania.

Pozostaje w tym miejscu zapytać, co dodatkowego wnosi analiza nieautomatyczna, wykraczająca poza kryteria strukturalne, polegająca na interpretowaniu poszczególnych przykładów i klasyfikowaniu ich ze względu na kryteria: kto, kogo, czym/pod jakim względem gorszy. W moim przekonaniu tak obrana ścieżka badawcza przede wszystkim pozwala dowiedzieć się więcej na temat tego, co gorszy ludzi: 15% wypowiedzi wskazuje na łamanie reguł obyczajowych (np. nadużywanie alkoholu/narkotyków, przekleństwa, awantury, tatuaże, życie z kimś bez ślubu), dodatkowo 14% zdań dotyczy nagości (w miejscach publicznych, w mediach, na zdjęciach, w teatrze itp., także rzeźb z widocznymi genitaliami), 11% to wypowiedzenia skoncentrowane wokół erotyki (pornografia, prostytutka, wyzywające/bezwstydne zachowania znanych osób itp.), 9% stanowią zdania, w których gorszące jest to, co szokuje drastycznością lub rażąco kontrastuje (np. zdjęcia na billboardach antyaborcyjnych, świętokradztwo, krzyże stawiane w nieodpowiednich miejscach); mniejsze grupy (poniżej 5%) obejmują przykłady, w których tym, co gorszy, są czyjeś poglądy (niezgodne z obowiązującymi w określonym środowisku), treści związane z szeroko pojętym nurtem gender oraz grupą osób LGBTQ+ lub epatowanie pieniędzmi. Wyodrębnione grupy dotyczą więc przede wszystkim norm obyczajowych i zasad społecznych – gorszy to, co niestosowne. Automatyczna analiza nie przyniosła tak szczegółowych wyników, pozwoliła jednak skutecznie wyodrębnić pewne kręgi tematyczne – ogółem liczba zdań odnoszących się w jakikolwiek sposób do treści związanych z katolicyzmem to aż 34% (z czego blisko 1/5 to wypowiedzi stanowiące parafrazę lub interpretację treści biblijnych), z kolei łączna liczba konkordancji dotyczących sztuki to 11%.

Grupę osób, które gorszą, należałoby nieznacznie poszerzyć – z analizy materiału wynika, że oprócz księży (10%) są to także politycy (7%). Samodzielna interpretacja przykładów nie wnosi jednak nowych informacji dotyczących grup osób, które są gorszone – z moich wyliczeń wynika, że są to kolejno: dzieci i młodzież (9%), wierni, parafianie, katolicy (3%), osoby określane rzeczownikiem *maluczcy* (3%) oraz te, których charakterystyka wynika z okoliczności, w jakich

się znajdują – widzowie, czytelnicy, przechodnie, świadkowie, naród itp. (każdy z tych zbiorów występuje w zdaniach poniżej 3%).

Samodzielna interpretacja przykładów niewątpliwie pozwoliła dostrzec więcej kontekstów, w których używany jest czasownik *gorszyć*, nieoceniony okazał się tu szeroki kontekst – nawet przy analizie nieautomatycznej pojedyncze zdanie z badanym wyrazem często nie wystarcza, by móc orzekać o jego znaczeniu. Metoda ta ma jednak swoje ujemne strony – może być stosowana w odniesieniu do niewielkich korpusów badawczych i naznaczona jest subiektywizmem (część zdań być może da się zinterpretować inaczej). Należy też zaznaczyć, że wybrany za przykład czasownik *gorszyć* pod pewnymi względami nie jest łatwy do analizy – reprezentuje grupę słownictwa nacechowanego aksjologicznie, którego definicje słownikowe odsyłają do imponderabiliów, a jego znaczenie i użycie nie jest takie samo w każdej społeczności w każdym czasie. Niezależnie od tych trudności zastosowanie wybranych funkcji automatycznej analizy języka przyniosło wiele istotnych wniosków, potwierdzonych podczas samodzielnego opracowywania materiału. Niewykluczone więc, że porównanie wyników badań nad czasownikiem z innego pola leksykalnego przyniosłoby jeszcze korzystniejszą (choć już teraz wysoką) ocenę narzędzi cyfrowych.

#### BIBLIOGRAFIA

##### SŁOWNIKI WRAZ ZE SKRÓTAMI

- ISJP – Bańko, M. (red.). (2000). *Inny słownik języka polskiego*. Wydawnictwo Naukowe PWN.
- SJPD – Doroszewski, W. (red.). (1958–1969). *Słownik języka polskiego*. Państwowe Wydawnictwo Naukowe.
- SL – Linde, S.B. (1807–1814). *Słownik języka polskiego*. Drukarnia XX. Pijarów.
- SPXVI – Mayenowa M.R., Peplowski, F. (red.). (1966–1994). *Słownik polszczyzny XVI wieku*. Ossolineum.
- Swar – Karłowicz, J., Kryński, A.A., Niedźwiedzki, W. (red.). (1900–1927). *Słownik języka polskiego*. Drukarnia E. Lubowskiego i S-ki.
- SWil – Zdanowicz, A. i in. (red.). (1861). *Słownik języka polskiego*. Maurycy Orgelbrand.
- WSJP PAN – Żmigrodzki, P. (red.). (2007-). *Wielki słownik języka polskiego*. Instytut Języka Polskiego PAN, <https://wsjp.pl>

##### LITERATURA

- Boryś, W. (2005). *Słownik etymologiczny języka polskiego*. Wydawnictwo Literackie.
- Duda, B., Przyklenk, J., Sujkowska-Sobisz, K. (2019). Reprezentacje humanistyki cyfrowej w polskim dyskursie naukowym. W: B. Ciesek-Ślizowska, B. Duda, E. Ficek, K. Sujkowska-Sobisz (red.). *Reprezentacje świata w dyskursach (modele, obrazy, wizje)* (ss. 55–69). Wydawnictwo Uniwersytetu Śląskiego.

- Eder, M. (2016). Słowa znaczące, słowa kluczowe, słowozbiory – o statystycznych metodach wyszukiwania wyrazów istotnych. *Przegląd Humanistyczny*, 3, 31–44.
- Grzegorzczkova, R. (2000). O znaczeniach współczesnych i ewolucji semantycznej wyrazów *gorszyć* (*się*), *zgorzyć* (*się*) i *zgorzenie*. *Prace Filologiczne*, 45, 215–220.
- Kieraś, A., Kobylński, Ł. (2021). Korpusomat – stan obecny i przyszłość projektu. *Język Polski*, 101 (2), 49–58.
- Marciszewski, W. (1977). *Metody analizy tekstu naukowego*. Państwowe Wydawnictwo Naukowe.
- Ossowska, M. (1957). Zagadnienie powszechnie uznanych norm moralnych. *Studia Filozoficzne*, 3, 80–96.
- Pawłowski, A. (2020). Lingwistyka kwantytatywna a humanistyka cyfrowa: ciągłość czy zmiana? W: A. Hącia, K. Kłosińska, P. Zbróg (red.). *Polszczyzna w dobie cyfryzacji* (ss. 239–253). Polska Akademia Nauk.
- Pawłowski, A., Piasecki, M., Broda, B. (2010). Możliwości i ograniczenia metody automatycznego generowania profili semantycznych leksemów na podstawie danych korpusowych. *Zeszyty Prasoznawcze*, 3–4 (203–204), 55–78.
- Pęzik, P. (2016). Exploring Phraseological Equivalence with Paralela. W: E. Gruszczyńska, A. Leńko-Szymańska (red.). *Polish-Language Parallel Corpora* (ss. 67–81). Instytut Lingwistyki Stosowanej UW.
- Pęzik, P. (2020). Budowa i zastosowania korpusu monitorującego MoncoPL. *Forum Lingwistyczne*, 7, 133–150.
- Piasecki, M. (2010). Automatyczne wydobywanie wiedzy o semantyce języka naturalnego z korpusów tekstu. W: P. Stalmaszczyk (red.). *Metodologie językoznawstwa. Filozoficzne i empiryczne problemy w analizie języka* (ss. 143–181). Wydawnictwo Uniwersytetu Łódzkiego.
- Przybylska, R. (2020). Kolokacje a analiza semantyczna wyrazu. *LingVaria*, 2(30), 43–51.
- Puzynina, J. (1982). Językoznawstwo a aksjologia. *Biuletyn Polskiego Towarzystwa Językoznawczego*, 39, 23–32.
- Radomski, A., Bomba, A. (red.). (2013). *Zwrot cyfrowy w humanistyce. Internet. Nowe media. Kultura 2.0*. E-naukowiec.

MOŻLIWOŚCI I OGRANICZENIA NARZĘDZI CYFROWYCH  
W BADANIACH NAD SŁOWNICTWEM NACECHOWANYM AKSJOLOGICZNIE  
(NA PRZYKŁADZIE CZASOWNIKA *GORSZYĆ*)

Streszczenie

Celem artykułu jest zestawienie wyników analiz przeprowadzanych przy użyciu narzędzi z zakresu lingwistyki cyfrowej z wnioskami wysnuwanymi w toku tradycyjnego opracowywania materiału. Przedmiotem badań jest nacechowany aksjologicznie czasownik *gorszyć*. Materiał wyekscerpowany z korpusu MoncoPL poddany został automatycznej analizie prowadzącej do pozyskania: kolokacji, listy frekwencyjnej użytych wyrazów, wykazu słownictwa kluczowego, zestawienia najczęstszych źródeł pochodzenia zdań, informacji o częstotliwości użycia badanego czasownika w poszczególnych latach; poddany został także modelowaniu tematycznemu. Uzyskane wyniki pozwoliły w pewnym stopniu odpowiedzieć na pytania: kto lub co, a także czym lub jak nas gorszy, jednak dopiero samodzielne opracowanie materiału przyniosło istotne doprecyzowania w tym zakresie.

**Słowa kluczowe:** lingwistyka cyfrowa; badania korpusowe; semantyka; *gorszyć*

POSSIBILITIES AND LIMITATIONS OF DIGITAL TOOLS  
IN RESEARCH ON AXIOLOGICALLY MARKED VOCABULARY  
(ON THE EXAMPLE OF THE VERB *GORSZYĆ*)

Summary

The aim of the article is to compare the results of analyzes carried out using tools in the field of digital linguistics with the conclusions drawn in the course of traditional material development. The subject of the research is the axiologically marked verb *gorszyć*. The material extracted from the MoncoPL corpus was subjected to automatic analysis leading to the acquisition of: collocations, a frequency list of the words used, a list of key words, a list of the most common sources of origin of sentences, information on the frequency of use of the verb in particular years; it was also subjected to thematic modeling. The obtained results allowed to some extent to answer the questions about the features of the use of the verb *gorszyć*, but only the independent development of the material brought significant clarifications in this regard.

**Keywords:** digital linguistics; corpora research; semantics; *gorszyć*