Adam Pawłowski
Uniwersytet Wrocławski
ORCID 0000-0002-0804-5698

LINGWISTYKA KWANTYTATYWNA A HUMANISTYKA CYFROWA: CIĄGŁOŚĆ CZY ZMIANA?

1. UWAGI WSTĘPNE: JAK W NAUCE ROZUMIEĆ KONTYNUACJĘ?

Modna ostatnio humanistyka cyfrowa (dalej HC) nie powstała *ex nihilo*. Czerpie ona z dorobku wcześniejszych badań humanistyki (w tym językoznawstwa), nauk społecznych i informatyki. Jednak jej związki z wypracowanymi dawniej teoriami i metodami, są złożone i nieoczywiste. Powstanie i rozwój HC oraz zmiany, zachodzące dziś w całej humanistyce, to nie łagodne wyłanianie się nowego paradygmatu badań, który zastępuje wcześniejszą perspektywę. Należałoby je uznać raczej za rewolucję, prowadzącą, przynajmniej na niektórych polach, do zerwania z wielowiekową tradycją refleksji humanistycznej. Dodatkowo proces ten jest powiązany ze zmianą pokoleniową, polegającą na stopniowej marginalizacji tych przedstawicieli humanistyki, których aparat metodologiczny został ukształtowany jeszcze w okresie gutenbergowskim, czyli do końca lat 80. ubiegłego wieku, a którzy z różnych powodów nie rozwinęli w dostatecznym stopniu kompetencji cyfrowych.

Na wstępie warto zastanowić się, jak w nauce można rozumieć kontynuację. Widziałbym tutaj dwa szerokie zakresy znaczeniowe. Pierwszy obejmowałby kontynuację faktyczną, opartą na kilku powtarzalnych schematach, drugi kontynuację „konstruowaną" przez potomnych bez podstawy materiałowej czy też bezpośrednich relacji konkretnych badaczy. Gdy mówimy o kontynuacji faktycznej, chyba najbardziej oczywistym przypadkiem jest istnienie „łańcucha edukacyjnego" nauczycieli i wychowanków (mistrzów i uczniów), którzy przekazują kolejnym pokoleniom wiedzę, doświadczenie i najlepsze praktyki. Takie rozumienie kontynuacji wydaje się oczywiste, ponieważ jest kalką sekwencji genealogicznej, której towarzyszy swoista transmisja kulturowa. W rzeczywistości jednak transmisja wiedzy naukowej, prowadzona mechanicznie i bezrefleksyjnie, może być wysoce problematyczna. Przekazywanie wiedzy i doświadczenia w nauce ma bowiem sens tylko wtedy, gdy owe wartości niematerialne są nowatorskie i dają kolejnym pokoleniom większe możliwości poznawcze. Jednak mikrosystemy społeczne (a szkoły czy grupy naukowe takowymi są) mają skłonność do kostnienia i zamykania się w kręgach idei, które jakoby wyznaczone zostały poglądami niegdysiejszych mistrzów. Można więc zadać pytanie: czy dzisiejsi leksykografowie, ilustrujący hasła przykładami czerpanymi z mowy potocz-

nej, znaleźliby uznanie w oczach tych, którzy dawniej sięgali praktycznie tylko po przykłady wzorcowego języka literatury? Czy autorzy wielkich dzieł leksykograficznych z przeszłości uznaliby za celowe publikowanie słowników wulgaryzmów i potocyzmów? Czy *Wielki słownik języka polskiego*, wiążący w przestrzeni cyfrowej dane korpusowe, opisy haseł i materiał multimedialny, mógłby opierać się na wzorach polskiej dziewiętnastowiecznej i dwudziestowiecznej leksykografii, która rozwijała się w zupełnie innej rzeczywistości? Otóż, mówiąc Asnykiem, nie da się rozwijać nauki, oglądając się nieustannie za siebie, powielając dawne wzorce. Warunkiem postępu nauki – oczywiście nie jedynym! – nie jest kontynuacja, lecz podważanie lekcji mistrzów wszędzie tam, gdzie można znaleźć lepsze rozwiązania problemów starych i gdzie pojawia się nowa rzeczywistość i nowy materiał do badań.

Kontynuacja w rozumieniu „łańcucha ludzkiego" ma więc sens ograniczony w czasie, dotyczyć też powinna kwestii zasadniczych, czyli etyki zawodowej uczonych lub podstaw metodologii. Oczywiście i te zagadnienia są dyskusyjne. Jak bowiem porównać metody *big data* z próbkowaniem na niewielką skalę? Jak odnieść się do uznawanej w świecie zasady powszechnego dostępu do wiedzy, skoro na przykład podręczniki retoryki i erystyki, tworzone z myślą o prowadzeniu szlachetnych kampanii społecznych, są wykorzystywane do niecnych celów w marketingu i propagandy? Jak oceniać szukanie popularności przez uczonych w mediach społecznościowych lub mieszanie nauki z polityką czy ideologią? Jak traktować odwieczny spór podejścia dedukcyjnego („nie muszę badać korpusów językowych, skoro potrafię *a priori* przewidzieć każde zdanie czy wypowiedzenie, niezależnie od tego, czy zostało atestowane w tekstach") z badaniem empirycznym i indukcją („badać można tylko to, co jest potwierdzone w danych korpusowych lub ankietach")? Uznając zasadę kontynuacji opartej na relacji mistrz–uczeń, należy więc mieć świadomość tego, że na większość pytań badawczych każde pokolenie musi poszukiwać odpowiedzi samodzielnie.

Drugi obszar kontynuacji faktycznej to konsekwentne i pogłębione wykorzystywanie literatury przedmiotu, niewymagające jednak bezpośrednich relacji mistrz–uczeń, jakie zachodzą między promotorami i doktorantami lub profesorami i ich podopiecznymi w tych samych ośrodkach. Takie podejście jest o wiele łatwiejsze, albowiem pozwala na niewielkie przeskoki w czasie i przestrzeni, niewymagające systematycznych kontaktów konkretnych osób lub pracy w tym samym miejscu. Ponadto daje ono możliwość wydobywania z czeluści magazynów bibliotecznych prac naukowych, które popadły w zapomnienie z powodów niemerytorycznych (praca w małym ośrodku, mało prestiżowy tytuł, pozorna kontrowersyjność proponowanych idei, skonfliktowanie autora ze środowiskiem).

Trzecim obszarem faktycznej kontynuacji badań jest trwałość jakiegoś zasobu. Przykłady takich źródeł to zwykle wielkie bibliografie, kartoteki słowników lub archiwalia (również w postaci nagrań lub materiałów fotograficznych), a w przypadku lingwistyki – także specyficzne zbiorowości ludzkie, które zachowują wartościowe z jakiegoś powodu cechy (na przykład dialektalne). Pod tym względem Polska, z przyczyn historycznych, jest podzielona, ponieważ tak zwane Ziemie Nowe, położone na zachodzie i północy kraju, nie odziedziczyły starych zasobów, powiązanych z kulturą narodową (wyjątkiem jest wrocławskie Ossolineum, posiadające niewielki ułamek przedwojennych zbiorów lwowskich), a archiwalia niemieckie cieszą się w Polsce ograniczonym zainteresowaniem. Także kształtujące się na zachodzie Polski

nowe dialekty mieszane nie były traktowane przez dziesięciolecia jako materiał szczególnie interesujący dla socjolingwistyki[1]. Natomiast ośrodki naukowe Warszawy, Krakowa oraz innych miast uniwersyteckich Polski centralnej mogą dziś rozwijać projekty digitalizacyjne i badawcze, oparte na własnych kartotekach czy archiwach, niejako kontynuując, w zupełnie nowych warunkach, dawne przedsięwzięcia naukowe z okresu międzywojnia i wcześniejszego (Polska Bibliografia Literacka, Nowy Korbut, różne projekty leksykograficzne itd.)[2].

Mówiąc o kontynuacji „konstruowanej", należy zwrócić uwagę na takie jej cechy, jak brak związków personalnych mistrz–uczeń, brak merytorycznych cytowań literatury przedmiotu oraz permanencję problemu badawczego (a nie zasobu materialnego jako przedmiotu badań). Przykładem takich ponadczasowych problemów są chociażby: sporne autorstwo tekstu, klasyfikacja tekstów, modelowanie struktur składniowych, analiza stylu czy rekonstrukcja struktur semantycznych. Nawiązania do wcześniejszej literatury, jeśli istnieją, mają charakter rytualny i powierzchowny, służący zadośćuczynieniu wymogom gatunku, ale nie odwołują się do treści cytowanych prac i nie polemizują z nimi.

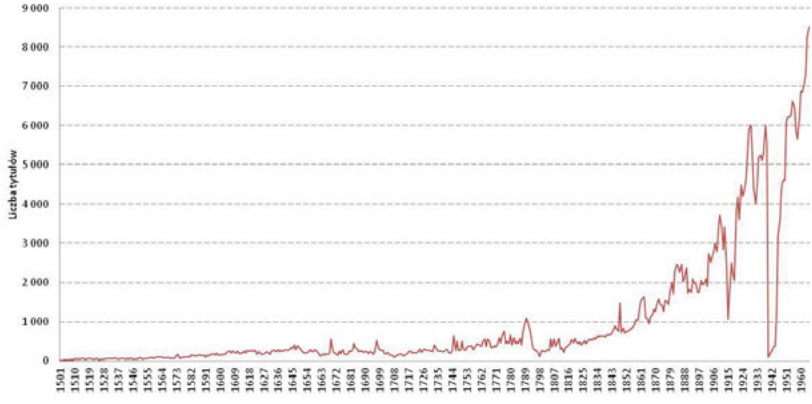
Istnienie jakiejś formy kontynuacji (lub jej brak) można też rozpatrywać na szerszym tle filozofii i historii nauki. Artykuł niniejszy nie dotyczy bezpośrednio kwestii epistemologii i ontologii, dlatego zasygnalizuję jedynie ważniejsze wątki takich rozważań. Jeżeli uczeni kontynuują istniejące wcześniej nurty badań (faktycznie lub wirtualnie), interesującym pytaniem jest zasięg czasowy takiej kontynuacji. W lingwistyce „czas życia" dzieł cytowanych merytorycznie w zasadzie nie przekracza 15–20 lat, wyjąwszy oczywiście analizy historyczne. Pytanie z dziedziny filozofii nauki dotyczy natomiast traktowania dłuższych serii zdarzeń, na przykład obejmujących stulecia, i odpowiedzi na pytanie, czy w nauce istnieją wielkie trendy historyczne (można nawiązać tutaj do popularnej wśród historyków teorii długiego trwania – por. Braudel 1958, 1969). W kontekście lingwistyki i humanistyki cyfrowej doszukujemy się na przykład trendu dezintegracyjnego, polegającego na tworzeniu się coraz węższych specjalności, lub integracyjnego, polegającego na agregacji sztucznie rozdzielanych tematów badań w szersze dziedziny (czego przykładem jest współcześnie właśnie humanistyka cyfrowa czy kognitywistyka, a wcześniej renesansowe rozumienie nauki). Można też analizować sposób, w jaki przebiega rozwój wiedzy (powolna akumulacja faktów i tworzenie teorii potwierdzających wcześniejsze poglądy lub rewolucja naukowa zmieniająca diametralnie i gwałtownie postrzeganie przedmiotu badań).

Wśród implikacji ontologicznych takich rozważań mieści się też pytanie o przyjętą koncepcję świata. Język można na przykład traktować jako byt dość chaotyczny, podlegający jednak procesom samoregulacji, można też przyjąć wizję teleologiczną, opierającą się na założeniu, że przynajmniej pewne parametry czy cechy języka mają stały kierunek rozwoju i „cel" (na przykład doskonalenie skuteczności komunikacji, obniżanie wysiłku artykulacyjnego itd.), można też akceptować bądź odrzucać istnie-

---

[1] Na przykład język Dolnego Śląska nie doczekał jeszcze pełnego, historycznego opisu, nie istnieje też repozytorium danych bardzo nielicznych badań terenowych po 1945 roku [por. Pawłowski, Tworek 2021].

[2] Osobną kwestią jest wykorzystanie obszernych zasobów badawczych, które z powodów historycznych przechowywane są poza dzisiejszymi granicami Polski.



Rysunek 1
Liczba tytułów książek wydawanych w Polsce w latach 1500–1965[3]

nie praw uniwersalnych, regulujących funkcjonowanie języka. Uwagi powyższe, choć z pozoru bardzo ogólne, znajdują liczne potwierdzenia w historii nauki o języku. Odnoszą się one także do relacji, jaka łączy lingwistykę kwantytatywną z humanistyką cyfrową. Brak zweryfikowanych danych naukometrycznych z zakresu językoznawstwa nie pozwala na ilustrację graficzną powyższych tez. Przedstawiam jednak histogram (rysunek 1) obrazujący wzrost liczby tytułów książek publikowanych w Polsce w latach 1500–1965, oparty na obszernej kwerendzie źródłowej [Czarnowska 1967]. Czy z poniższego wykresu można wnioskować, że mamy do czynienia z wielowiekowym procesem historycznym? A jeżeli tak, to czy można będzie odkryć podobne trendy (procesy?) w obszarze humanistyki i lingwistyki? Chociaż granica między faktografią a interpretacją danych jest w podobnych przypadkach trudna do uchwycenia, kwestie kontynuacji, zerwań lub nagłych przewrotów w różnych dyscyplinach nauki można opisywać taką właśnie metodą modelowania naukometrycznego.

2. BADANIA ILOŚCIOWE JĘZYKA

Wyczerpujący opis podejścia ilościowego do języka wymagałby w zasadzie osobnej, obszernej, monografii. Tutaj przedstawię jedynie podstawowe założenia i najważniejsze obszary problemowe takich badań. Warto przypomnieć, że najpełniejszy dotychczas opis dorobku polskiej lingwistyki kwantytatywnej (dalej LK), autorstwa niżej podpisanego i prof. Jadwigi Sambor, ukazał się w pracy *Quantitative linguistics in Poland* [Pawłowski, Sambor 2005]. Wydawnictwo *Quantitative Linguistik / Quantitative Linguistics. Ein Internationales Handbuch / An International Handbook*, z którego pochodzi cytowana praca, można też polecić czytelnikom zainteresowanym historią i dorobkiem LK, ponieważ zamieszczono tam rozdziały na temat badań prowadzonych w większości państw świata oraz liczne artykuły problemowe o charakterze

[3] Dane pochodzą z pracy [Czarnowska 1967]. Wiele danych lingwistycznych dla polszczyzny, częściowo ilustrujących teorię długiego trwania, zawiera praca [Górski et al. 2019].

encyklopedycznym. Poniżej przedstawiam podstawowe założenia podejścia ilościowego w nauce o języku. Są one na tyle ogólne, że zachowują aktualność zarówno w odniesieniu do badań dotychczasowych, jak i przyszłych. Następnie krótko omawiam główne obszary badań lingwistyki kwantytatywnej w Polsce, stosując metodę modelowania tematycznego na korpusie bibliografii prac z zakresu LK.

Cechy podejścia ilościowego można streścić w kilku punktach. Podstawą jakichkolwiek prac czy rozważań ilościowych jest uznanie języka za strukturę lub sieć segmentowalną, a więc złożoną z elementów dyskretnych, powiązanych ze sobą i tworzących różne hierarchie. Drugim istotnym założeniem badań ilościowych jest mierzalność i kwantyfikowalność elementów wyróżnionych na różnych poziomach granulacji tekstu (od głoski po obszerne, ale semantycznie spójne ciągi zdań). Z obu tych fundamentalnych założeń łatwo wysnuć wniosek, że tekst (także w postaci korpusu) jako najważniejsza i prymarnie jakościowa manifestacja języka może mieć różne reprezentacje numeryczne. Ich liczba zależy tylko od tego, jak wiele zastosuje się metod kwantyfikacji jednostek językowych, czyli w praktyce najczęściej głosek, morfemów, sylab, wyrazów, fraz i zdań. Kolejne założenie badań ilościowych stwierdza, że chociaż nie wszystkie reprezentacje numeryczne tekstu są relewantne z punktu widzenia konkretnego celu badań, niektóre można za takowe uznać. Oznacza to, że reprezentacje niedające się interpretować w kategoriach lingwistycznych lub nieefektywne pod względem poznawczym, są odrzucane. Gdy powyższe warunki są spełnione, a więc gdy istnieją relewantne reprezentacje numeryczne tekstu traktowanego jako byt prymarnie jakościowy, przyjmujemy, że możliwe jest przetwarzanie tych reprezentacji metodami statystyki i matematyki. Ponieważ jednak matematyka dysponuje ogromnym arsenałem metod i z liczbami jest w stanie zrobić praktycznie wszystko, istotne jest tylko takie przetwarzanie danych tekstowych, które daje wyniki interpretowalne w kategoriach lingwistycznych i poznawczo wartościowe. I chociaż zabrzmi to jak truizm, w kontekście wielu prac ilościowych o języku, pisanych przez fizyków czy matematyków bez wykształcenia językoznawczego, warunkiem wstępnym uznania ich dzieł za „poznawczo wartościowe" jest zrozumiałość. Sztuka prowadzenia badań ilościowych nie polega bowiem jedynie na szybkim liczeniu, lecz na właściwym doborze hipotez, analizowanych jednostek, sposobów ich kwantyfikacji i, dopiero w ostatniej fazie, właściwych algorytmów.

Warto zwrócić jeszcze uwagę na inną cechę badań ilościowych języka, niezwiązaną bezpośrednio z samą procedurą przetwarzania danych. Otóż metody ilościowe nie są przypisane wyłącznie do lingwistyki. Wielokrotnie okazywały się one przydatne w badaniach literaturoznawczych, czego najlepszym przykładem jest badanie autorstwa dzieł literackich, filiacja tekstów (powiązanie na osi czasu różnych wersji tego samego dzieła – na przykład „Bogurodzicy"), statystyczna analiza stylu, ale także leksykalne profilowanie utworów i/lub pisarzy (tutaj punktem wyjścia jest zawsze wygenerowanie listy frekwencyjnej). Podobne spostrzeżenie dotyczy socjologii i politologii: analizy frekwencyjne różnych tekstów użytkowych (niemal wyłącznie na poziomie leksykalnym) są często wykorzystywane w badaniach prowadzonych w ramach tych dyscyplin. Można też zastanowić się na kwestią podobieństwa założeń lingwistyki kwantytatywnej i strukturalizmu. Istotnie, lingwistyka kwantytatywna posługuje się pojęciami jednostek, poziomów analizy, hierarchii i dystrybucji. Jednak łączenie przyczynowo-skutkowe obu tych nurtów badań byłoby nadużyciem. Badania ilościowe prowadzo-

no już w XIX wieku, zanim narodził się strukturalizm, prowadzi się je również obecnie, w kontekście innych teorii.

Kwestia wskazania listy dominujących tematów lingwistyki kwantytatywnej jest o tyle kłopotliwa, że zakres wykorzystania metod ilościowych jest dziś praktycznie nieograniczony. Można jednak wskazać tematy specyficzne dla tego kierunku, w zasadzie niereprezentowane w innych nurtach badań języka. W tej grupie znalazłoby się badanie statystycznych praw językowych, regulujących funkcjonowanie dowolnego etnolektu. Tradycyjnie wymienia się tutaj prawa Zipfa i Menzeratha-Altmanna, ale także Kryłowa czy Beöthy [Hammerl, Sambor 1997; Kułacka 2011]. Należy zauważyć, że prawa te mają zastosowanie wykraczające daleko poza granice lingwistyki i są często opisywane w ramach ogólnej teorii systemów, gdzie wykorzystuje się nie tylko dane lingwistyczne, lecz także socjologiczne, ekonomiczne czy demograficzne. Innym, specyficznym obszarem badań ilościowych są dociekania spornego autorstwa, a szerzej – taksonomie automatyczne tekstów literackich i użytkowych. Bardzo stabilnym i dynamicznym obszarem badawczym jest leksykografia komputerowa, która przy okazji różnorodnych badań wykorzystuje na coraz większą skalę aparat metod ilościowych. Warto też wspomnieć o badaniach z zakresu „poetyki ilościowej", prowadzonych w latach 60. głównie przez pracowników i współpracowników Instytutu Badań Literackich PAN [por. Mayenowa 1965].

Obecnie, dzięki postępom technologicznym i badaniom wielkich korpusów, metody ilościowe są coraz częściej wykorzystywane do modelowania struktur semantycznych (*nota bene* właśnie w niniejszej pracy zastosowałem jedną z nich, a mianowicie *topic modeling*, czyli modelowanie tematyczne). 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. Można dodać, że w początkach dwudziestego wieku fakt ten przyczynił się do powstania i rozwoju nurtu behawioralnego w lingwistyce, programowo odrzucającego spekulacje na temat tego, „co poeta miał na myśli". Rozwój technologii i informatyki sprawił jednak, że narzędzia służące do empirycznych badań semantycznej warstwy języka wreszcie powstały. Szczególnie istotne są tutaj dwa obszary. W neurolingwistyce można stosować obserwować i rejestrować procesy zachodzące w mózgu człowieka podczas aktywności językowej, opierając się na technice pozytonowej emisji tomograficznej (PET – *positron emission tomography*). Technika ta jest w zasadzie nieinwazyjna i oprócz najbardziej rozpowszechnionych zastosowań medycznych służy lingwistyce i logopedii.

Z punktu widzenia lingwistyki kwantytatywnej istotniejszy był jednak inny obszar. Dzięki powstaniu wielkich korpusów tekstów i algorytmom sztucznej inteligencji udało się stworzyć narzędzia skutecznie profilujące pod względem semantycznym konkretne leksemy, a więc i całe teksty. W szczególności chodzi o analizy nastawienia tekstu (ang. *sentiment analysis*)[4], generowanie różnego rodzaju relacji między leksemami oraz

[4] Termin angielski *sentiment analysis* jest tłumaczony przez środowisko inżynierów języka jako *analiza wydźwięku*. Jest to niezbyt udana adaptacja, ponieważ *wydźwięk* zakłada wiedzę o recepcji tekstu, do której analiza automatyczna nie ma dostępu (wydźwięk jest *dla kogoś*, nie istnieje więc przy braku odbiorcy). Dlatego proponuję mówić o analizie nastawienia

systemy dialogowe, pozwalające algorytmom sztucznej inteligencji na komunikację z człowiekiem (tak zwane chatboty). Rozwój Internetu i zalew informacji sprawiły, że pojawiło się zapotrzebowanie na takie badania, co pozwoliło na sfinansowanie interdyscyplinarnych zespołów, tworzących coraz doskonalsze programy analizy automatycznej. O ile więc dawne problemy lingwistyki kwantytatywnej nie zniknęły i dalej prowadzi się intensywne badania ilościowe „wielkich tekstów kultury" lub weryfikuje prawa językowe, o tyle szeroki strumień finansów, skierowany na rozwój badań semantyki, sprawił, że inaczej rozłożone są dziś akcenty w obrębie całej dyscypliny.

Wracając do historii badań ilościowych w Polsce, można w syntetyczny sposób przedstawić rozkład tematów zainteresowań tej dyscypliny w Polsce, dodając, iż nie odbiega on w istotny sposób od analogicznego rozkładu w świecie. Materiałem empirycznym, który pozwolił na wygenerowanie przedstawionych poniżej wyników, była bibliografia polskiej lingwistyki kwantytatywnej do roku 2000. Liczy ona ok. 600 pozycji, głównie w języku polskim. Analizie poddano tytuły traktowane jako mikroteksty, przy czym wersje obcojęzyczne były podawane w tłumaczeniu polskim. Celem badania było wskazanie relewantnych i reprezentatywnych dla całego nurtu LK wyrazów oraz zbitek wyrazowych, które można określić mianem słów kluczowych. Poniższe zestawienia (tabele 1 i 2) zawierają wyniki, których interpretacja nie powinna stanowić problemu. Wprawdzie brak jest nadrzędnych terminów klasyfikacyjnych, takich jak *fonetyka*, *morfologia* czy *semantyka*, jednak obecne są inne terminy, wskazujące na zainteresowanie polszczyzną mówioną (*mowa*, *widmo mowy polskiej*), historią języka (*nieregularny rozwój fonetyczny spowodowany frekwencją*), ale przede wszystkim leksyką. Jednak cechą specyficzną LK, widoczną w obu listach słów klu-

Tabela 1

Słowa kluczowe wygenerowane z bibliografii polskiej lingwistyki kwantytatywnej w układzie rangowym[5]

[język]	[struktura]	[materiał]
[tekst]	[mowa]	[poezja]
[polszczyzna]	[lingwistyka]	[statystyka]
[badanie]	[rozkład]	[forma]
[analiza]	[wiersz]	[korpus]
[słownictwo]	[rozwój]	[językoznawstwo]
[wyraz]	[styl]	[frekwencja]
[metoda]	[częstość]	[długość]
[słownik]		

tekstu, które jest pochodną stosunku nadawcy do przedmiotu wypowiedzenia. Nastawienie jest wyrażone konkretnymi środkami językowymi i podlega analizie empirycznej.

5 Do wygenerowania danych z tabel 1 i 2 zastosowano narzędzie TermoPl, udostępnione przez CLARIN-PL (https://ws.clarin-pl.eu/termopl.shtml).

Tabela 2
Zbitki wyrazowe wygenerowane z bibliografii polskiej lingwistyki kwantytatywnej w układzie rangowym

[język polski]	[ujęcie statystyczne]	[materiał słownika frekwencyjnego]
[współczesny język polski]	[nieregularny rozwój fonetyczny spowodowany]	[długość sylabiczna wyrazów]
[słownik frekwencyjny]	[tekst]	[zastosowanie metod statystycznych]
[słownictwo współczesnego języka polskiego]	[statystyczne badanie cech osobniczych głosu]	[materiał Pana Tadeusza]
[współczesna polszczyzna]	[metoda statystyczna]	[problem atrybucji tekstu]
[polszczyzna mówiona]	[prawo ilościowe rozwoju języka]	[materiał języka polskiego]
[lista frekwencyjna]	[polski język ogólny XIX]	[styl współczesnej polszczyzny]
[język]	[polszczyzna]	[prawdopodobieństwo subiektywne wyrazów]
[lingwistyka kwantytatywna]	[zastosowanie metody]	[średni rozkład energii]
[analiza statystyczna]	[badanie]	[widmo mowy polskiej]

czowych, jest nacisk na metodę, a nie materiał czy poziom języka. Istotnie, poza oczywistymi słowami kluczowymi *język* i *językoznawstwo* to właśnie *metoda* pojawia się najczęściej. Swoistym komentarzem do przedstawionych tutaj wyników mogą być prace *Słowa i liczby* (Sambor 1972) oraz *Statystyka dla językoznawców* [Hammerl, Sambor 1990], zawierające dyskursywne opisy zagadnień tej dyscypliny w drugiej połowie dwudziestego wieku.

Modelowanie tematyczne pozwoliło natomiast na wyróżnienie kilku słowozbiorów (topików), wskazujących na następujące obszary badań: analiza tekstu (1), słownictwo (2), rozkłady częstości wyrazów (3), struktura języka (4), analiza mowy (5), metody statystyczne (6), stylometria (7) i słownik frekwencyjny (8). Topik 7 jest o tyle interesujący, że zawiera terminy związane z analizą statystyczną metryki wiersza – dziś w zasadzie niepraktykowaną. Słowozbiory przedstawiam poniżej w kolejności odpowiadającej podanym liczbom[6].

Podsumowując powyższe rozważania, należy powiedzieć, że nurt lingwistyki kwantytatywnej mógł w pełni rozwinąć się w łonie strukturalizmu, nie będąc jednak jego częścią. Profil tych badań był (i jest) empiryczny, nastawiony na metody, a nie konkretne typy zjawisk. W przeszłości dominowało badanie formy, a nie treści języka,

[6] Topiki wygenerowano metodą LDA jednym z narzędzi CLARIN-PL http://ws.clarin-pl.eu/topic.shtml. Przykład innych badań z zastosowaniem tej techniki opisano w pracy [Pawłowski, Walkowiak 2018].

Rysunek 1
Słowozbiory odpowiadające obszarom badań ilościowych języka

obecnie warstwa semantyczna jest już także eksplorowana. Do słabych stron LK należy przesadny scjentyzm, nieadekwatność metod i hermetyczność tej dyscypliny, wynikająca ze stosowania aparatu matematycznego. Nowszymi wcieleniami nurtu LK stały się lingwistyka korpusowa (jej popularność jednak słabnie, ponieważ badania korpusowe weszły w obręb humanistyki cyfrowej) oraz właśnie humanistyka cyfrowa.

3. HUMANISTYKA CYFROWA: PRÓBA DEFINICJI

Zakres terminu „humanistyka" jest, przynajmniej pozornie, określony dość jasno. Według *Słownika języka polskiego* PWN pod red. Mieczysława Szymczaka chodzi albo o „nauki badające człowieka jako istotę społeczną oraz jego wytwory, na przykład język, literaturę, sztukę itd.", albo o wydziały uczelni wyższych, na których takowe nauki są uprawiane. Inne kompendia leksykograficzne w zasadzie niczego nowego do tej definicji nie wnoszą. Na przykład *Wielki słownik języka polskiego* (WSJP) dodaje jeszcze znaczenie trzecie („kierunek studiów"), ściśle powiązane z drugim, samą zaś definicję formułuje dość podobnie: „ogół nauk innych niż nauki przyrodnicze i techniczne, stawiających w centrum zainteresowania człowieka, jego język, twórczość i funkcjonowanie w społeczeństwie"[7]. Przytoczone tutaj definicje w jakimś stopniu sankcjonują istniejące podziały, częściowo je ustalają, ale w praktyce pozostawiają wiele luk i niejasności, ponieważ są li tylko konwencjami podporządkowanymi doraźnym potrzebom administrowania instytucjami nauki. Tę krytyczną opinię pośrednio potwierdza formuła definicji zawartej w WSJP. Jest ona oparta na logicznym dopełnieniu zbioru („inne niż..."), a nie definiowaniu „pozytywnym" („takie jak..."), co stanowi jakiś rodzaj kapitulacji w obliczu konieczności uwzględnienia wielkiej różnorodności tego, co jest „inne niż nauki przyrodnicze i techniczne", ma „w centrum zainteresowania człowieka, jego język, twórczość i funkcjonowanie w społeczeństwie" i na ogół (chociaż nie zawsze) samoidentyfikuje się jako humanistyka.

Można zatem przyjąć, że jeżeli powyższe definicje dobrze wyznaczają zakres pojęcia humanistyki, cyfrowość będzie jedynie jej rozszerzeniem o nowe metody i formy bądź formaty zapisu na różnych nośnikach. Jeżeli jednak definicje te potraktujemy jako niespójne i ogólne konwencje, służące głównie temu, by utrwalać porządek administracyjny nauki instytucjonalnej, w której jednostki badawczo-dydaktyczne powinny mieć przypisane wyraziste profile i specjalności tylko po to, by podlegać innym jednostkom o szerszych uprawnieniach – humanistyka cyfrowa będzie mieć kłopot z własną tożsamością. Dotychczasowa praktyka badań cyfrowych różnych wytworów kultury duchowej człowieka – transdyscyplinarnych i najczęściej prowadzonych zespołowo – zdaje się potwierdzać to drugie założenie. Nie jest na przykład jasne, kiedy jakiś „wytwór człowieka" jako istoty społecznej staje się przedmiotem zainteresowania humanistyki. Co sprawia, że książka takowym przedmiotem czasem jest (literatura piękna), a czasem nie jest (część piśmiennictwa użytkowego)? Kiedy zwykły produkt przemysłowy (na przykład puszka z zupą) przechodzi przez fazę transcendencji, by stać się przedmiotem kultury? Czy wystarczy określić go mianem wytworu inte-

[7] http://www.wsjp.pl/

lektu lub narysować i wyeksponować w stosownym kontekście, dodając ciekawą narrację? Osobnej refleksji wymaga też język, stanowiący dotychczas centralną kategorię humanistyki. Jest on przecież jednocześnie materialny i abstrakcyjny, biologiczny i kulturowy, indywidualny i społeczny. Owe aspekty, czy też fasety, języka są ze sobą tak blisko powiązane, że ich odseparowanie jest praktycznie niemożliwe. Wątpliwości idą jeszcze dalej, gdy za kategorię centralną humanistyki uzna się człowieka – a sugerują to nie tylko wszystkie definicje, lecz także nadal przejrzysta etymologia nazwy dyscypliny (łac. *homo*, czyli „człowiek"). Przecież czynności intelektualne, polegające na podejmowaniu decyzji bądź przetwarzaniu informacji (automatyczne tłumaczenie, konwersacja, generowanie tekstu lub obrazu), mogą dziś być wykonywane przez programy sztucznej inteligencji i konstruowane przez inżynierów humanoidy. Można zatem zapytać, domeną jakiej dyscypliny będzie komputerowe przetwarzanie tekstu wygenerowanego automatycznie przez inny komputer. Wracając do cytowanych na wstępie definicji humanistyki, należałoby więc może powiedzieć, że współcześnie jej elementem centralnym jest człowiek bądź jego sztuczne awatary symulujące kompetencję komunikacyjną.

Uwzględniając powyższe uwagi, humanistykę cyfrową (HC) definiuję jako: zbiór praktyk poznawczych oraz zasobów wiedzy i informacji, będących efektem przeniesienia do sfery cyfrowej i twórczego rozszerzenia (ilościowego i jakościowego) praktyk poznawczych humanistyki opartej na druku. Można dodać, że jest HC przede wszystkim działaniem (*techne*), a mniej wiedzą (*doxa*, *episteme*), kładzie nacisk na infrastruktury i potencjał poznawczy, a nie na gotowe odpowiedzi i rozwiązania problemów. Podejście, jakie proponuję, odrzuca też mechanicznie ujmowaną opozycję tradycja *vs* nowoczesność i niskie wartościowanie humanistyki gutenbergowskiej. Tym, co charakteryzuje prace w obszarze HC i odróżnia ją od praktyk gutenbergowskich, jest istnienie „łańcucha produkcji", czyli ciągu czynności prowadzących do wytworzenia „produktu" w postaci infrastruktury badawczej. Łańcuch taki rozpoczyna się od wygenerowania zapisu elektronicznego z danych źródłowych i jego anotacji, prowadzi dalej do stworzenia koncepcji architektury informacyjnej systemu, interfejsu komunikacyjnego, aż do w pełni działającego systemu[8]. Ponadto humanistyka cyfrowa odróżnia się od tradycyjnej w warstwach metodologicznej i społeczno-komunikacyjnej.

W perspektywie metodologicznej HC charakteryzuje się wieloma cechami, spośród których najważniejsze są: praca na wielkich zbiorach danych (na przykład korpusach tekstów); filozofia Linked Open Data, polegająca na integracji i rozbudowie powiązanych metadanymi wielkich zbiorów danych w różnych formatach (tekst, głos, obraz i geolokalizacja); bezprecedensowa w historii nauki rola metadanych i grafiki; uwzględnianie czynnika komunikacyjnego, czyli stałego dostępu użytkowników i łatwych transmisji danych; uwzględnianie aspektów prawnych, związanych z prawami autorskimi.

Humanistyka cyfrowa odróżnia się też wieloma aspektami praktycznymi (można określić je jako społeczno-komunikacyjne). Wśród nich należałoby wymienić takie cechy, jak: kolektywny charakter pracy (konsorcja, zespoły badawcze), zadaniowość

---

[8] Rozważania na temat humanistyki cyfrowej znaleźć można m.in. w pracach [Radomski, Bomba 2013]; [Maryl 2014]; [Schreibman et al. 2004]; [Vitali-Rosati, Sinatra 2014].

(realizacja projektów), międzynarodowy zasięg, przewaga komunikacji sieciowej nad bezpośrednią, wykorzystanie wolontariuszy i społecznościowych form współpracy (na przykład crowdsourcing) czy hiperteksowość (a nie linearność) produktów HC.

4. CZY MOŻNA MÓWIĆ O KONTYNUACJI BADAŃ ILOŚCIOWYCH W OBSZARZE HUMANISTYKI CYFROWEJ?

Zanim przejdę do krótkiego porównania zakresu humanistyki cyfrowej i lingwistyki kwantytatywnej, przytoczę trzy przykłady badań ilościowych, które mimo swojej wartości nie były często cytowane w okresie najbujniejszego rozwoju tej dyscypliny i do dziś w zasadzie pozostają na marginesie lingwistyki polskiej. Mówiąc o początkach leksykologii kwantytatywnej, nie sposób nie wymienić dzieła autorstwa Fabiana Ferdynanda Sławińskiego (1830–1903) *Obliczenie wyrazów zawarte w trzech słownikach: 1) Lindego 2) w Wileńskim 3) Rykaczewskiego* [Sławiński 1873]. Ta zaiste prorocza i do dziś wartościowa praca, zawierająca m.in. obliczenia przydatne w stenografii (odpowiednik dzisiejszej kompresji danych) i leksykologii kwantytatywnej (rozkład frekwencyjny liczby znaczeń na poszczególne leksemy trzech słowników), nie była cytowana i wykorzystywana jako symboliczny początek, a zarazem inspiracja do nowych dociekań, ani w pracach nad *Słownikiem frekwencyjnym polszczyzny współczesnej*, ani też nad Narodowym Korpusem Języka Polskiego. Kontynuacji badań więc nie było.

W czasie, gdy w oczach badaczy młodego pokolenia to firma Google odkrywa możliwości analizy kultury i języka na podstawie frekwencji leksemów (ang. *culturomics*), warto przytoczyć wypowiedź Fabiana Ferdynanda Sławińskiego – praktycznie nieznanego lingwisty z nieistniejącego w dziewiętnastym wieku państwa: „Na zamknięcie niniejszej pracy powiem: Jeżeli zoologowie uważają za pożyteczną wiadomość policzenie ziarn ikry różnych ryb; jeżeli ziemianie sporządzają księgi i inwentarz, w którym spisują pilnie swój dobytek: dla czegóż wyrazy, wyrób myśli, inwentarz językowy, niezasługiwały u nas na odpowiednie uwzględnienie? [...] Liczby, wyrażające ogólny zbiór wyrazów niewieleby pouczały, ale rozklasyfikowane choćby tak, jak w niniejszej pracy podałem, mają cechę wyższego znaczenia" [Sławiński 1873, 36]. Niemal 150 lat po tej publikacji badacze doszukują się właśnie owego „wyższego znaczenia" w materiale tekstowym, ekscerpowanym w ogromnych ilościach bezpośrednio ze źródeł (czyli bez przetworzenia leksykograficznego), posługując się w tym celu algorytmami AI, które uczą się reguł działania od człowieka.

Drugi przykład historycznych antecedencji badań współczesnych dotyczy analizy zawartości, a pośrednio także nastawienia. Podejście to zastosował w nauce polskiej jako pierwszy pedagog krakowski Jan Władysław Dawid (1859–1914). W pracy *O zarazie moralnej: studyum psychologiczno-społeczne* z 1886 roku, opisującej m.in. niekorzystne skutki zamieszczania opisów przemocy w prasie codziennej [Dawid 1886], badacz ten przedstawił dokładne wyliczenie klas zjawisk pozytywnych i negatywnych w obrębie „wiadomości z charakterem moralnym", wyróżniając m.in. „czyny dobre", „opisy podniecające żądzę posiadania", „czyny złe bez kary", „czyny złe ukarane" itd. Co ciekawe, obliczenia te, stanowiące kwintesencję współczesnej analizy zawartości i nastawienia („sentiment analysis"), zawarte zostały na zaledwie jednej stronie

w formie tabelki, natomiast reszta książki poświecona została rozważaniom pedagogicznym i moralnym [Dawid 1886, 212]. W tym przypadku osoba autora została jednak przywrócona współczesności dzięki sumienności prof. Walerego Pisarka, który przed wielu laty wprowadził analizę zawartości do badań prasoznawczych, przypominając metodę pracy z tekstem, stworzoną sto lat wcześniej przez Dawida [Pisarek 1983]. Można więc powiedzieć, że częściowa kontynuacja badań rzeczywiście miała miejsce, chociaż profil naukowy obu badaczy był zupełnie różny, dzieliło ich także całe stulecie.

Przykład trzeci ma związek z badaniami stylometrycznymi, których inicjatorem i jednym z pionierów był filozof Wincenty Lutosławski (1863–1954). Stworzył on termin *stylometria* i przeprowadził jedne z pierwszych w historii badań tego typu [Lutosławski 1896, 1897a, 1897b]. Niestety ze względu na dość niekonwencjonalny charakter, silnie narodową orientację polityczną i poglądy uważane przez część środowiska naukowego za kontrowersyjne, jego osiągnięcia nie były wśród lingwistów okresu PRL znane i doceniane. Stan ten zaczął powoli ulegać zmianie po tym, jak opublikowałem serię prac poświęconych lingwistycznym aspektom jego spuścizny [Pawłowski 2004a, 2004b, 2006; Pawłowski, Pacewicz 2004; Pawłowski, Zaborowski 2006]. Chociaż o kontynuacji faktycznej jego myśli mówić dziś trudno, szczególnie w kontekście nowych technik obliczeniowych, przedstawione przed ponad wiekiem przez Lutosławskiego założenia ogólne stylometrii, w tym analizy glottochronologicznej (ustalenie chronologii dzieł Platona na podstawie ewolucji cech stylistycznych), znajdują współcześnie pełne potwierdzenie.

Powyższe przykłady pokazują, że o stabilnym rozwoju badań ilościowych do okresu poprzedzającego powstanie humanistyki cyfrowej można mówić tylko w odosobnionych przypadkach. Co ciekawe, okres Drugiej Rzeczpospolitej nie pozostawił w tym zakresie wielkich osiągnięć – poza kilkoma przypadkami[9] prac ilościowych praktycznie wówczas nie prowadzono. Natomiast przeskok od badań języka, prowadzonych w okresie rozbiorowym, do nauki w czasach PRL, a więc w państwie obciążonym piętnem ideologii i częściowo odciętym od nauki światowej, okazał się niewykonalny.

Przejście od lingwistyki kwantytatywnej do humanistyki cyfrowej przebiegało natomiast w zupełnie innych warunkach. Miało charakter dość płynny, a fakt, że rewolucja technologiczna zbiegła się ze zmianą systemu politycznego w 1989 roku, nie miał aż takiego znaczenia. Wśród cech wspólnych obu nurtów należy wymienić przede wszystkim: skrajny empiryzm (indukcję, a nie dedukcję); wiarę w to, że w nauce (a więc i w językoznawstwie) jakość wiedzy jest prostą pochodną ilości danych; uznanie metod ilościowych za skuteczne; prymat metody nad problemem. Poza tym w humanistyce cyfrowej kontynuowane (chociaż w innym zakresie) są prace ilościowe nad warstwą leksykalną, popularna jest też atrybucja tekstów i analizy stylometryczne.

Jednak różnic i zerwań jest znacznie więcej, tak iż dla wielu młodych adeptów HC lingwistyka kwantytatywna pozostaje obszarem słabo znanym. Przede wszystkim jednak HC w obecnej odsłonie jest domeną bardzo szeroką, integrującą (z lepszym lub

---

[9] Statystykę wykorzystywał m.in. Witold Doroszewski [por. Doroszewski 1938]. Artykuł na ten temat (zapewne zainspirowany młodzieńczymi pracami Wincentego Lutosławskiego, które recenzował jako profesor Uniwersytetu w Dorpacie) napisał także Jan Baudouin de Courtenay [1927].

gorszym skutkiem) wiele odległych od siebie dyscyplin, obejmujących nie tylko humanistykę, lecz także nauki społeczne (medioznawstwo, politologię, socjologię), a nawet archeologię. Konferencje humanistyki cyfrowej przypominają dziś odkładane od dawna zjazdy rodzinne, na których spotykają się dawno niewidziani krewni, o których istnieniu już zapomnieliśmy, a na których widok uświadamiamy sobie, iż mimo funkcjonowania w różnych instytucjach coś nas jednak łączy. Socjolog spotyka tam na przykład lingwistów, dowiadując się ze zdziwieniem, że potrafią oni doskonale analizować dane tekstowe z ankiet, lingwista z kolei odkrywa badania muzykologów nad systemami zapisu dźwięków i może od siebie dodać, że zajmuje się tym także semiotyka, historyk uczy się od informatyków anotacji danych, a od geografów tworzenia map itd. Jest więc współczesna humanistyka cyfrowa inter- lub transdyscyplinarna, otwarta na użytkownika (portale różnych projektów zamiast tradycyjnych publikacji), jej utylitaryzm jest ważniejszy od funkcji poznawczej, jest też w pewnym sensie arogancka i nieświadoma swojego pochodzenia, ponieważ niechętnie odwołuje się do prac z okresu przedcyfrowego.

Jeśli wśród tych różnic należałoby wskazać tę, która dla językoznawstwa jest szczególnie istotna, byłaby nią pewna degradacja języka jako podstawowego kodu modelującego całokształt procesów poznawczych (tak przynajmniej uważano w dwudziestym wieku) i sprowadzenie go do roli narzędzia. Ponadto dyscypliną integrującą różne nurty humanistyki cyfrowej stała się nie filozofia jako matka wszystkich nauk, lecz informatyka, która jako jedyna jest w HC wszechobecna. Ujmując to nieco przewrotnie, dzisiejsza humanistyka cyfrowa jest wyjątkowo niehumanistyczna. Czy stan ten uznać należy za permanentny? Oczywiście nie, zapewne kiedyś pojawi się inna forma współistnienia informatyki i dyscyplin utożsamiających się z humanistyką cyfrową. Wracając jednak na podwórko lingwistyczne, należy pogodzić się z faktem, iż nie da się już powstrzymać wpływu technologii informacyjnych język, polegającego na masowym korzystaniu z systemów sztucznej inteligencji, symulującej kompetencję językową człowieka.

LITERATURA

Baudouin de Courtenay J., 1927, Ilościowość w myśleniu językowym [w:] Symbolae gramaticae in honor J. Rozwadowski, Vol. I, Kraków, 3–18. Przedruk [w:] J. Baudouin de Courtenay, 1990, Dzieła wybrane, t. IV, Warszawa, s. 546–563.

Braudel F., 1958, Histoire et Sciences sociales: La longue durée, „Annales. Économies, Sociétés, Civilisations", nr 4 (1958), s. 725–753.

Braudel F., 1969, Écrits sur l'histoire, Paris.

Czarnowska M., 1967, Ilościowy rozwój polskiego ruchu wydawniczego 1501–1965: dane szczegółowe o książkach 1929–1938 i 1951–1960 oraz o czasopismach 1933–1937 i 1956–1938, Warszawa.

Dawid J.W., 1886, O zarazie moralnej: studyum psychologiczno-społeczne, Warszawa.

Doroszewski W., 1935, Pour une représentation statistique des isoglosses, „Bulletin de la Société Linguistique de Paris", 36, p. 28–42.

Górski R., Król M., Eder M., 2019, Zmiana w języku. Studia kwantytatywno-korpusowe, Kraków.

Hammerl R., Sambor J., 1990, Statystyka dla językoznawców, Warszawa.
Hammerl R., Sambor J., 1993, O statystycznych prawach językowych, Warszawa.
Kułacka A., 2011, Statystyczne prawa językowe: na przykładzie prawa Menzeratha-Altmanna w składni języków polskiego i angielskiego, Wrocław.
Lutosławski W., 1896, Sur une nouvelle méthode pour déterminer la chronologie des dialogues de Platon (mémoire lu le 16 mai 1896 à l'Institut de France), Paris.
Lutosławski W., 1897a, On stylometry. Abstract of a paper read at the Oxford Philological Society on May 21st by Dr. W. Lutosławski, of Drozdowo, near Lomza, Poland, „Classical Review" 11, 1897, p. 284–286.
Lutosławski W., 1897b, The origin and growth of Plato's logic, London, New York, (reprint: Hildersheim: Georg Olms Verlag, 1983).
Maryl M., 2014, Odświeżanie filologii, „Teksty Drugie" 146 (2), s. 9–20.
Mayenowa M.R., red., 1965, Poetyka i matematyka, Warszawa.
Pawłowski A., 2004a, Wincenty Lutosławski – a forgotten father of stylometry, „Glottometrics" 8, s. 83–89.
Pawłowski A., 2004b, Travail de maîtrise de Wincenty Lutosławski: «Aesthetisches Studium. Ueber das phonetische Element in der Poesie». Description matérielle et analyse du contenu, „Organon" 33, s. 121–139.
Pawłowski A., 2006, Wincentego Lutosławskiego „Aesthetisches Studium. Ueber das phonetische Element in der Poesie". Opis formalny i analiza merytoryczna, „Kwartalnik Historii Nauki i Techniki" 51(2), s. 47–68.
Pawłowski A., Pacewicz A., 2004, Wincenty Lutosławski (1863–1954): Philosophe, helléniste ou fondateur sous-estimé de la stylométrie, „Historiographia Linguistica" 31 (2/3), s. 423–447.
Pawłowski A., Sambor J., 2005, Quantitative linguistics in Poland [w:] Quantitative Linguistik / Quantitative Linguistics. Ein Internationales Handbuch / An International Handbook, red. R. Köhler, G. Altmann, R. Piotrowski, Berlin, New York, s. 115–130.
Pawłowski A., Tworek A., 2021, Dolny Śląsk, Polska, Europa. Studium komunikacji, Wrocław (w druku).
Pawłowski A., Walkowiak T., 2018, Topic modelling as a Tool for Researching the Polish Daily Press Corpus ChronoPress of the Post-war Period (1945–1962). Proceedings of the 8th Conference of Japanese Association for Digital Humanities, Tokyo, September 9–11, p. 27–29, https://conf2018.jadh.org/files/Proceedings_JADH2018_rev0911.pdf.
Pawłowski A., Zaborowski R., red., 2006, Wincenty Lutosławski – oblicza różnorodności, Drozdowo.
Pisarek W., 1983, Analiza zawartości prasy, Kraków.
Radomski A., Bomba R. (red.), 2013, Zwrot cyfrowy w humanistyce Internet – Nowe Media – Kultura 2.0, Lublin.
Schreibman S., Siemens R., Unsworth J. (red.), 2004, A Companion to Digital Humanities, Oxford.
Sambor J., 1972, Słowa i liczby. Zagadnienia językoznawstwa statystycznego, Wrocław, Warszawa etc.
Sławiński F.F., 1873, Obliczenie wyrazów zawarte w trzech słownikach: 1) Lindego 2) w Wileńskim 3) Rykaczewskiego, Warszawa.
Vitali-Rosati M., Sinatra M.E. (red.), 2014, Pratiques de l'édition numérique, Montréal.