


Magdalena Pastuch 
Uniwersytet Śląski, Katowice
magdalena.pastuch@us.edu.pl

Kinga Wąsińska 
Uniwersytet Śląski, Katowice
kinga.wasinska@us.edu.pl

Barbara Mitrenga 
Uniwersytet Śląski, Katowice
barbara.mitrenga@us.edu.pl

ANOTACJA SOCJOLINGWISTYCZNA W KORPUSIE DAWNYCH POLSKICH TEKSTÓW DRAMATYCZNYCH (1772–1939)¹

Słowa kluczowe: historia języka polskiego, lingwistyka korpusowa, historyczna pragmalingwistyka, socjopragmatyka, anotacja socjolingwistyczna

Keywords: history of the Polish language, corpus linguistics, historical pragmatology, sociopragmatics, sociolinguistic annotation

1. Wprowadzenie

Lingwistyka korpusowa ma obecnie silną, ustabilizowaną pozycję w obszarze badań językoznawczych. Na przestrzeni kilku już dekad jej rozwoju wyraźnie można zaobserwować nie tylko doskonalenie się tej metodologii, ale także zmiany, jakie dokonały się w rozumieniu podstawowego dla niej pojęcia, czyli korpusu. Najbardziej ogólne, słownikowe znaczenie zakłada, że jest to duży zbiór tekstów stworzony według ściśle określonych kryteriów. W badaniach korpusowych wskazuje się na

¹ Publikacja dofinansowana ze środków przyznanych w ramach programu Inicjatywa Doskonałości Badawczej Uniwersytetu Śląskiego w Katowicach. Wkład autorek w powstanie tekstu jest taki sam i wynosi 33,33%.

dodatkowe dwie właściwości: reprezentatywność i zrównoważenie (Pęzik 2013). Kolejną istotną cechą korpusu jest anotacja zawartego w nim materiału językowego, której typ zależy od celów, jakim ma służyć tworzony korpus.

Założeniem artykułu nie jest przedstawienie kierunków rozwoju językoznawstwa korpusowego w Polsce, opis istniejących korpusów czy sposobów ich wykorzystania. Na wstępie warto jedynie zwrócić uwagę na szybki rozwój lingwistyki cyfrowej, wielonurtowość prowadzonych prac korpusowych oraz zróżnicowanie tworzonych baz danych². W dalszej części szczegółowemu opisowi poddano jeden z aspektów tworzenia korpusów – anotację tworzących go tekstów. Dzięki zastosowaniu określonego typu anotacji automatycznych – morfologicznej, morfoskładniowej (Woliński, Przepiórkowski 2001; Hajnicz, Murzynowski, Woliński 2008) – bądź zaimplementowaniu anotacji wielopoziomowych, to jest gramatycznej oraz socjolingwistycznej czy pragmatycznej³, zwiększa się zakres użycia wyszukiwarki, a tym samym dostępność zdeponowanego materiału językowego. W artykule przedstawiono zasady anotacji socjolingwistycznej wprowadzonej w opracowanym w Instytucie Językoznawstwa Uniwersytetu Śląskiego w Katowicach we współpracy z konsorcjum CLARIN-PL Korpusie dawnych polskich tekstów dramatycznych (1772–1939) (KorTeDa)⁴, którego upublicznienie planowane jest w 2024 r.

Jest to pierwszy w Polsce anotowany socjolingwistycznie korpus 50 dawnych tekstów dialogicznych⁵, o objętości około 900 tysięcy tokenów, który znajdzie zastosowanie w historycznych badaniach pragmalingwistycznych rozwijanych od ponad 20 lat w językoznawstwie światowym, a w Polsce obecnych ciągle jeszcze w nielicznych pracach (np. Cybulski 2003; Pawłowska 2014). W artykule przedstawiono przyjęte kryteria oraz zasady anotacji socjolingwistycznej, natomiast charakterystyka korpusu, metadanych, wykorzystanego narzędzia Inforex (zob. Marcińczuk, Kocoń, Broda 2012) oraz szczegółowa instrukcja używania wyszukiwarki korpusowej będą przedmiotem odrębnej publikacji. W zakończeniu artykułu pokazano przykładowe zastosowanie KorTeDa w historycznych badaniach pragmalingwistycznych ze świadomością, że korpus ten może być wykorzystany również w badaniach literaturoznawczych, historycznych i innych.

-
- 2 Syntetyczne omówienie projektów cyfrowych poświęconych dawnej polszczyźnie oraz podsumowanie osiągnięć humanistyki cyfrowej w Polsce w zakresie słowników (ich digitalizacji, tworzenia wersji elektronicznych, opracowania narzędzi do przeszukiwania danych słownikowych) zob. Pastuch, Duda, Liszyk et al. 2018. Informacje o elektronicznych zasobach językowych gromadzi przewodnik opracowany przez Patrycję Pałkę i Agatę Kwaśnicką-Janowicz (2017). Lista zasobów cyfrowych znajduje się na stronie: <http://clip.ipipan.waw.pl/LRT>.
 - 3 Przegląd istniejących schematów anotacji pragmatycznej przedstawiono w artykule Dawn Archer, Jonathana Culpepera i Matthew Daviesa (2008).
 - 4 Koncepcja korpusu została szczegółowo omówiona podczas posiedzenia Komisji Historii Języka Polskiej Akademii Nauk 23 XI 2023 r. Zarys projektu przedstawiono w: Mitrenga, Pastuch, Wąsińska 2021.
 - 5 Kryteria doboru tekstów w KorTeDa opisano w: Mitrenga, Pastuch, Wąsińska 2021.

2. Socjopragmatyka w badaniach historycznojęzycznych

Prowadzone przez nas badania mają profil historyczno-pragmatyczny, a ich podstawowym celem jest (re)konstrukcja dawnej polszczyzny potocznej. Są one osadzone w dwóch, bliskich sobie pod wieloma względami, paradygmatach metodologicznych: (mikro)pragmatyce⁶ i socjopragmatyce. Omówienie kryteriów wprowadzonej anotacji wymaga krótkiego przedstawienia założeń historycznych badań socjopragmatycznych. W najogólniejszym rozumieniu jest to ta część pragmalingwistyki, która koncentruje się na czynnikach socjologicznych, bliska jest więc socjolingwistyce. Należy jednak pamiętać, że socjopragmatyka poza wskazywaniem socjologicznie uwarunkowanych modeli komunikacyjnych zajmuje się także objaśnianiem, jak te utrwalone wzorce są używane w konkretnych sytuacjach komunikacyjnych (zob. Culpeper 2009: 3). Nieco inaczej kształtują się zadania socjopragmatyki, jeśli odniesiemy się do materiału historycznego. Choć od czasów przełomowego artykułu Andreea Juckera (1994) historyczna pragmalingwistyka (w tym socjopragmalingwistyka) rozwija się bardzo intensywnie, to jednak z powodu ograniczeń materiałowych (coraz częściej z powodzeniem przełamanych) ma ona inny profil niż badania synchroniczne. Bazowanie na tekstach zapisanych, pierwotnie mówionych, lub na dialogach konstruowanych⁷ powoduje, że analizy te bliższe są pragmatologii. Podstawowym celem jest opis zależności pomiędzy użyciem języka a kontekstem historyczno-społecznym, a następnie wykorzystanie do ich objaśnienia stosownych narzędzi pragmalingwistycznych.

Nasze badania odnoszą się zatem do trzech subdyscyplin: językoznawstwa korpusowego, historii języka i pragmalingwistyki. Wskazane obszary mają wiele miejsc wspólnych, ale na różnych etapach realizacji projektu zakłada się różną ich hierarchię. W przedstawianym artykule opisano etap, na którym uwaga skoncentrowana była na tworzeniu narzędzia, którym jest wspomniany wcześniej korpus. Ze względu na przyjętą metodologię badawczą oraz cel prowadzonych badań, konieczne było zastosowanie, poza nałożoną automatycznie anotacją morfologiczną, anotacji socjolingwistycznej, czyli takiej, która uwzględni podstawowe parametry socjologiczne,

6 Anotacja umożliwiająca badania w tym nurcie została w KorTeDa przeprowadzona ręcznie i obejmuje dziewięć kategorii tekstowych (m.in. formy grzecznościowe, interiekcje, przerywniki leksykalne, powtórzenia). Ten poziom anotacji podporządkowany jest badaniom dawnej polszczyzny potocznej prowadzonym przez autorki korpusu.

7 Za dialogi konstruowane uważa się w diachronicznych badaniach pragmalingwistycznych dialogi wyekscerpowane z tekstów dramatycznych i prozatorskich, które uznaje się za podstawę badań, mimo że nie są to dialogi autentyczne (zapisane). Stanowią one zbiór mieszczący się w tzw. *speech-related texts*, do których zalicza się zarówno mowę zapisaną (np. protokoły sądowe, protokoły sejmowe, zeznania i relacje świadków), jak i mowę konstruowaną (np. dramat, partie dialogowe w prozie, podręczniki do nauki języka i rozmówki) (Culpeper, Kytö 2000: 176). W obecnej wersji KorTeDa zawarte są wyłącznie dialogi konstruowane.

jakimi są wiek, płeć i status społeczny. W KorTeDa odnoszą się one do wypowiedzi postaci dramatów.

W największym historycznym korpusie angielskich tekstów dialogowych A Corpus of English Dialogues 1560–1760 (CED) (Kytö, Walker 2006) wydzielono część anotowaną socjopragmatycznie⁸, ale z co najmniej dwóch powodów nie można było w pełni wzorować się na rozwiązaniach wypracowanych na potrzeby języka angielskiego. Po pierwsze, korpus dialogów angielskich zawiera bardziej zróżnicowane, a tym samym bogatsze informacyjnie teksty. Są to nie tylko dramaty, ale także zapiski sądowe, które umożliwiają na przykład określenie roli zawodowej uczestnika procesu sądowego. Dodatkowo wskazuje się w tej części CED, obejmującej lata 1640–1760, relacje między nadawcą a odbiorcą wypowiedzi. Autorzy nazywają tę anotację socjopragmatyczną, a stosowną część korpusu nazwali The Sociopragmatic Corpus (SPC). Drugi powód, istotniejszy, wynika z faktu, że tego typu anotacja jest ściśle związana z realiami społeczno-historycznymi, tak odmiennymi dla XVII- i XVIII-wiecznej Anglii (Archer, Culpeper 2003). Zaświadcza ją chociażby bardziej szczegółowe wartości statusu społecznego⁹. Należy też zaznaczyć, że w KorTeDa anotowane są tylko tak zwane wartości statyczne, to znaczy te, które przypisane są na stałe do danej postaci, zaś w SPC oznacza się także wartości dynamiczne, to znaczy związane ze zmiennością ról społecznych danej postaci.

Bliskie nam rozumienie anotacji przedstawił wiele lat temu Geoffrey Leech, który zdefiniował ją jako „the practice of adding interpretative, linguistic information to an electronic corpus of spoken and / or written language data” (Leech 1997: 2)¹⁰. Wydaje się, że w przypadku anotacji socjolingwistycznej ten czynnik interpretacyjny jest szczególnie ważny.

3. Anotacja socjolingwistyczna w KorTeDa

Korpus zakłada oznaczenie wszystkich partii dialogowych i monologowych (poza zbiorowymi) trzema zmiennymi socjolingwistycznymi. Na podstawie informacji dostępnych w spisie postaci, zamieszczonych w didaskaliach oraz dających wyczytać się z treści dramatu, każda postać została scharakteryzowana pod względem wieku, płci i statusu społecznego. Dąży się, by w korpusie jak najmniej wypowiedzi miało status nieokreślony, stąd wyłożone starania w znalezieniu argumentów dla przypisania konkretnej wartości dla każdej z kategorii. Wykorzystuje się w tym celu

8 Korpus obejmuje lata 1560–1760, a część anotowana socjopragmatycznie to lata 1640–1760.

9 W warstwie średniej ówczesnego społeczeństwa angielskiego można było, ze względu na inną dynamikę rozwoju uwarunkowaną m.in. rozwojem przemysłu, wskazać bardziej szczegółowe podziały. W cytowanym artykule zawarta jest bogata bibliografia z zakresu historii społecznej Anglii. Ze względów oczywistych nie może być ona wykorzystana w odniesieniu do polskich realiów.

10 Za: Archer, Culpeper 2003: 5.

różne informacje, to znaczy: ważną rolę odgrywa akcja dramatu; zwraca się uwagę na to, jak w dramacie postrzegana jest postać przez pozostałych bohaterów; wysnuwane są wnioski na podstawie sposobu zwracania się do konkretnej osoby oraz tego, jak postać sama postrzega siebie i wypowiada się o sobie; analizowane są koneksje między postaciami. Do oceny włączane są także dane, które można wyczytać z tekstu na temat wyglądu fizycznego postaci i jej temperamentu.

Wszystkie te informacje zostały umieszczone w przygotowanych formatkach, na podstawie których poszczególnym postaciom dramatu zostały przypisane wartości anotacyjne. Niejednokrotnie formatki były rozbudowane, co wynikało z treści i konstrukcji dramatu.

Rys. 1. Fragment formatki dramatu (Sz)

lp.	Postać z opisem	Nazwa postaci w tekście głównym	Płeć	Status społeczny	Wiek
1.	HRABIA JAN. przyjaciel Doktora Karola; krewny Kassylidy; oficer jazdy polskiej.	HRABIA	mężczyzna	wysoki	dorosły średni
2.	WŁADYSŁAW. przyjaciel Doktora Karola; oficer jazdy polskiej.	WŁADYSŁAW	mężczyzna	średni	dorosły średni
3.	DOKTOR KAROL.	DOKTOR	mężczyzna	średni	dorosły średni
		JEZDZIEC	mężczyzna	średni	dorosły średni
4.	KASSYLDA. majątek; wysoki status materialny	KASSYLDA	kobieta	wysoki	dorosły młody
5.	EUSTACHY. brat Kassylidy. majątek; wysoki status materialny	GŁOS MĘZKI	mężczyzna	wysoki	dorosły młody
		EUSTACHY	mężczyzna	wysoki	dorosły młody
6.	ROMUALD. aptekarz.	ROMUALD	mężczyzna	średni	dorosły średni
7.	HELENA. córka aptekarza; Jakob do niej mówi Panniczka	HELENA	kobieta	średni	dorosły młody
8.	JAKOB. służący Doktora Karola; od 7 lat służy u Karola	JAKOB	mężczyzna	niski	dorosły średni
9.	PIERWSZY LEKARZ	PIERWSZY LEKARZ	mężczyzna	średni	dorosły średni
10.	DRUGI LEKARZ	DRUGI LEKARZ	mężczyzna	średni	dorosły średni

Źródło: opracowanie własne.

Rysunek 1 przedstawia fragment formatki dramatu (Sz)¹¹, w którym uczestniczą 22 postaci. Niektóre z nich mają więcej niż jedną nazwę, co wynika z intrygi i przebiegu wydarzeń, na przykład Karol występuje jako Doktor oraz Jeździec; nieznany na początku Głos Mężki (sic!) przypisano do Eustachego. W innej sztuce (Sr) młody syn wojewody, pułkownik powracający z dziesięcioletniego wojażowania wciela się w siedem różnych osób, które różni status społeczny i wiek. Walery, udający między innymi hrabiego, bankiera, ubożego chłopaka i poczciwego staruszka, w ostatniej scenie wyznaje: *Pod różnemi postaciami badałem serca wasze.*

11 Wszystkie przytoczone w artykule cytaty pochodzą z KorTeDa. W spisie źródeł podano rozwinięcie stosowanych w artykule skrótów tytułów dramatów.

Przypisanie odpowiednich wartości w formacie wymagało każdorazowo namysłu, tak aby wszystkie sygnały tekstowe zostały prawidłowo zinterpretowane i oznaczone stosownymi wartościami w obrębie kategorii anotacyjnych. Przygotowane formatki posłużyły do półautomatycznego oznakowania każdej indywidualnej wypowiedzi postaci.

Rys. 2. Widok anotacji socjolingwistycznej aktu V, sceny II dramatu (A) z wykorzystaniem narzędzia Inforex

AKT V. SCENA II

GUWERNER: (czyta bilety) **Tak jest**

XIAŻĘ LONGVILLE: **Mamy to szczęście być do Domu jego adresowanemi | oto**

PANI FORESTER: **To straszna rzecz | a na jakimże okręcie Panowie płynęli ?**

an#12511957:annotation mezczyzna
 an#12511956:annotation dorosly_sredni
 an#12511958:annotation sredni
 an#12511963:annotation mezczyzna
 an#12511962:annotation dorosly_senior
 an#12511964:annotation wysoki
 an#12511978:annotation kobieta
 an#12511979:annotation sredni
 an#12511977:annotation dorosly_sredni

Źródło: opracowanie własne.

Anotowany w ten sposób korpus daje możliwość przefiltrowania zgromadzonego materiału i wyświetlenia konkretnych wyników, które interesują badacza, na przykład samych wypowiedzi kobiecych, niezależnych od płci wypowiedzi osób o statusie niskim lub tylko wypowiedzi mężczyzn w średnim wieku o wysokim statusie społecznym. W systemie można dodatkowo zawęzić pole obserwacji do przedziałów czasowych lub miejsca wydania dramatów. Daje to możliwość badania zespołu nawyków mownych charakteryzowanych w kategoriach idiolektu (Klemensiewicz 1961; Walczak 2017), ale również pozwala na analizy lingwistyczne zjawisk wspólnych, niezależnych od indywidualnych cech językowych autorów dramatów.

3.1. Wiek

3.1.1. Rozumienie kategorii

Choć mogłoby się wydawać, że podanie w spisie postaci dokładnego wieku metrykalnego postaci, na przykład w formie: *Wojciech Migdał 25 lat* (K), *Cecylia 27 lat* (Pm), *Ksawera 38 lat* (Ka), *Major 54 lata* (Dh), pozwala na określenie wieku z pełną precyzją, warto pamiętać, że liczba lat człowieka nie przekłada się jednoznacznie na jego wiek biologiczny, psychologiczny oraz społeczny (Zaśko-Zielińska 2009).

Historycy podają, że w wieku XVIII średnia długość życia człowieka wynosiła nieco ponad 30 lat, a wiek później wydłużyła się o około 10 lat (zob. Kukło 2009; Kukło, Łukasiewicz, Leszczyńska 2014). W tym kontekście wiek metrykalny jawi się jako wartość względna.

Brak stałości i jednoznaczności w literaturze przedmiotu w sposobie definiowania wieku wymusiły uwzględnienie w tej kategorii dodatkowych założeń. W korpusie podział wypowiedzi ze względu na wiek oparty został na ustaleniach dotyczących faz życia jednostki, które najczęściej ujmowane są w czterech etapach: dzieciństwo, młodość, wiek średni i starość (Dobrowolska 1992: 79). Największy wpływ na wskazanie czasu dzieciństwa mają czynniki biologiczne, które później ustępują czynnikom społeczno-obyczajowo-kulturowym. Współcześnie decydującym kryterium przy określaniu danej grupy wiekowej jako młodzież jest sytuacja edukacyjna. Zakłada się, że granicą pomiędzy wiekiem młodzieńczym a dorosłością jest zakończenie etapu kształcenia się i podjęcie pracy zarobkowej. Jednak dla okresu, z którego pochodzą teksty korpusowe, to kryterium nie jest dystynktywne. Za dorosłe uznaje się osoby, które rozpoczęły czas przybierania największej liczby różnych ról społecznych. W charakterystyce języka ludzi starszych szczególnie przydatne są natomiast obserwacje czynników psychologicznych i somatycznych (np. luki w pamięci, problemy ze słuchem lub z poruszaniem się, zaburzenia logicznej wypowiedzi).

Wymienione fazy życia jednostki posłużyły za ramy kategorii wiekowych stosowanych w korpusie. Dopasowano do nich charakterystykę wieku postaci z uwzględnieniem czasu od jej narodzin, stanu rozwoju zdrowia w danym momencie życia oraz liczby pełnionych przez postać funkcji społecznych.

3.1.2. Zasady oznaczania wieku w KorTeDa

Kategoria wieku jest w korpusie ograniczona do sześciu możliwości: niedorośli, dorośli młody, dorośli średni, dorośli senior, dorośli nieokreślony, nieokreślony. Ogólny opis poszczególnych kategorii przedstawia się następująco:

- a) **niedorośli** – znacznik przypisywany jest wypowiedziom dzieci. Są to zazwyczaj synowie lub córki, którzy w utworze są najmłodszy. Często mają zdrobniałą formę imienia, na przykład *Helenka*, *Olenio*, *Walerek*, *Ignas*, którą może dopełniać określenie *mały/mała*, na przykład *Mała Jadzia* – córka Heleny i Bartnickiego (Ż), *Mały Antek* – sześciolatek syn Doktora i Maryi (Wd). W spisie postaci niekiedy notuje się wiek metrykalny dzieci, na przykład *Karolek lat 5* oraz *Franciszek lat 13* – synowie szlachcica Rzetelnickiego (Po); *Giguś Thugut 9-letni syn Thugutowej* (CM). Postaci niedorośle określane są w tekście na przykład jako: *dziecko*, *chłopiec*, *córka*.
- b) **dorośli młody** – znacznik obejmuje wypowiedzi postaci, które reprezentują najmłodsze pokolenie dorosłych. Są to młode panny szykujące się do zamąż-

pójścia, młodzi kawalerowie starający się o żonę, czyli synowie i córki na wydaniu albo nowożeńcy. Nie wskazujemy konkretnej granicy dorosłości, mając na uwadze zmieniającą się cezurę adolescencji, pierwotnie wyższą dla chłopców niż dla dziewcząt. Różne były też interpretacje pełnoletności: inna związana z możliwością zawarcia związku małżeńskiego, a inna – z uzyskaniem praw politycznych. Kobiety, nawet po osiągnięciu tak zwanych lat dojrzałych, nie nabywały zdolności do niektórych czynności prawnych. Ich zależność od mężczyzn trwała nawet mimo osiągnięcia wieku dojrzałego, a gasła dopiero w momencie zamążpójścia lub wstąpienia do klasztoru (Bednaruk 2019). Z uwagi na niski wiek zawierania związków małżeńskich mylnie mogą być etykiety typu *wdowa*, *kawaler*, *panna*. W tekście dramatu postaci takie określane są jako „młodzi”, na przykład *młoda pokojówka*, *młoda wdowa*.

Informację o wieku postaci należy potwierdzić uwagami w didaskaliach lub tekście, na przykład Stebelska nazywa Karskiego *młokosem* (Zż); Laura, spoglądając w oczy Kalenickiego, *patrzy w źrenice młodzieńca* (K); Roman Prysznicki chwali się świetną kondycją fizyczną: *nie mogę porzucić gimnastyki, łyżwowania, fechtunku, konnej jazdy* (PI), Jan – masztalerz właściciela dworu – goni konie, skacze (R). Młodzi dorośli to osoby wykonujące określone zawody (służące, pokojówki, parobczaki) lub postaci często określane jako *kolega*, *kuzynka*, *bratanek* głównej młodej postaci dorosłej.

c) **dorośli średni** – w ten sposób zaznaczane są wypowiedzi postaci, które reprezentują pokolenie średnie. To najczęściej rodzice, matka, ojciec, wuj, stryj, kuzyni rodziców, osoby w tak zwanym kwiecie wieku, a także podstarzałe panny, guwernantki, niemłode służące i nianie, starsi kawalerowie, rozwodnicy, na przykład Kajcio mówi o Ksawerze: *trochę przestarzała; jakaś stara kobieta, to pewno ciotka córki pana Hipolita*; ojciec Ksawery – Hipolit wyznaje: *teraz córkę za mąż wydawać trzeba, a wydają ją już chwala Bogu lat dwadzieścia* (Ka).

Wśród postaci drugoplanowych często jako osoby w średnim wieku określane są w didaskaliach lub tekście gospodarze, właściciele wsi, też literaci, urzędnicy oraz osoby od wielu lat służące w kościele. Postaci kobiece w średnim wieku zazwyczaj występują pod nazwiskami odmężowskimi, na przykład *Ogonowa*, *Maniewiczowa*, *Walentowa*, lub tytułuje się je funkcją męża, na przykład *Pułkownikowa*, *Mecenasowa*. Regularnie występujące zależności można zauważyć w stosowaniu nazw typu *Szambelanowa* – żona szambelana (dorośli średni), *Pani Jowialska* – matka szambelana (dorośli senior).

d) **dorośli senior** – w ten sposób oznaczane są wypowiedzi osób dorosłych, które reprezentują najstarsze pokolenie. Są to najczęściej dziadkowie, babki, nestorzy rodu, starcy, staruszki, emeryci wojskowi, na przykład Radost w rozmowie z Anielą mówi: *już szósty krzyżyk kończę* (Wg). W dramatach są to postaci określane jako „stare”, czasem z charakterystycznymi cechami starości (mające siwe włosy, zmarszczki, problemy ze wzrokiem, słuchem, a nawet z poruszaniem się),

na przykład Edward Karski mówi o swojej ciotce słowami *starowina* (Zż); Pan Łowicz-Mączewski opisany jest jako *bezczelnie łysy, bardzo dystyngowany stary epuzer, apoplektyczny kark, szkła na rasowym nosie* (Dok). Seniorzy pojawiają się również wśród służby i lokajów, jednak zawsze z informacją, że posługują od lat w danym domu. Mówią o sobie: *stara niania, staruszek, stary dziad, stary sługa starego bohatera* itp.

- e) **dorosły nieokreślony** – zdarza się, że na podstawie informacji zawartych w daskaliach oraz w treści dramatu wiadomo, że postać to osoba dorosła, ale nie można doprecyzować, czy jest to dorosły młody, średni lub senior. Sytuacja taka dotyczy postaci pobocznych, które wypowiadają się rzadko, czasem na przestrzeni całej sztuki zaledwie jednym zdaniem. Postaci dorosłe nieokreślone występują w tekście głównie pod nazwami: *kobieta, pracownica, krewny, tancerz, urzędnik, stolarz, dozorca, listonosz, gość, ordynans porucznika*.
- f) **nieokreślony** – w obecnej wersji korpusu znacznik ten nie został wykorzystany. Zakładamy jednak, że może być przydatny w sytuacji, gdy żadne z zastosowanych kryteriów nie umożliwi identyfikacji pewnych głosów, okrzyków czy szeptów z konkretną postacią. W obecnej wersji korpusu zdarzyło się to jedynie w wypowiedziach zbiorowych, ale te nie podlegają omawianej anotacji socjolingwistycznej.

Stworzone wytyczne pozwoliły uzyskać możliwie największą spójność ustaleń dla tekstów pochodzących z różnych okresów, co przełożyło się na operacyjność danych korpusowych i możliwość grupowania wypowiedzi pod względem wspólnego wieku postaci. Można zauważyć, że akcja wielu dramatów rozgrywa się w obecności przedstawicieli kilku pokoleń: dziadków, rodziców i dzieci – starszych, które są uwikłane w wątki miłosne, oraz młodszych, które jako niesamodzielne wymagają opieki i troski. Niekiedy są to dzieci młodych dorosłych. Dobrze strukturę wieku przedstawia sztuka (PJ). Przedstawicielami najstarszego pokolenia są Pan Józef Jowialski i Pani Małgorzata Jowialska, dorosłych średnich reprezentują Jan i Barbara Jowialscy (tytułowani również nazwą urzędu, który pełni mąż-ojciec, jako Szambelan i Szambelanowa), a dorosłych młodych – ich córka Helena, zakochana w Januszu.

3.2. Płeć

3.2.1. Rozumienie kategorii

Przy wskazaniu płci nadawców wypowiedzi przyjęto jej rozumienie w kategoriach płci biologicznej (ang. *sex*), a nie społeczno-kulturowej (ang. *gender*) (Szacka 2011). Decyzja o wyborze takiej perspektywy wynikała z przekonania, że odwołanie się do tradycyjnego podziału na postaci kobiece oraz męskie będzie rozwiązaniem odpowiednim dla badanego okresu historycznego (1772–1939), poprzedzającego rewolucję

w postrzeganiu płci w kategoriach społeczno-kulturowych oraz rozwój od lat 70. ubiegłego wieku *gender studies* (Slany, Struzik, Wojnicka red. 2011; Karwatowska, Szpyra-Kozłowska 2012). Wszystkie dramaty wchodzące w skład KorTeDa powstały w czasie, gdy jeszcze nie rozwinęła się refleksja nad płcią społeczno-kulturową, przyjęcie na etapie anotacji rozróżnienia nadawców wypowiedzi na podstawie płci biologicznej uwzględnia intencje autorów i autorek dramatów, którzy poszczególne partie mówione „włożyli w usta” kobiet lub mężczyzn.

Obecnie obszarem szczególnego zainteresowania lingwistów jest problematyka relacji między językiem a płcią, o czym wspomina między innymi Artur Rejter, zwracając uwagę, że „od pewnego czasu uczeni skupiają się coraz częściej na płci pojmowanej w kategoriach kulturowych (*gender*) niż biologicznych (*sex*)” (Rejter 2013: 28)¹². Mimo że na etapie anotacji nie rozróżnia się płci społeczno-kulturowej, na podstawie zastosowanego w KorTeDa podziału postaci zgodnie z ich płcią biologiczną i przypisaniu wypowiedziom wartości: kobieta lub mężczyzna¹³, możliwe jest pogrupowanie wypowiedzi na kobiece i męskie, co daje szerokie możliwości podjęcia szczegółowych badań z zakresu języka kobiet i mężczyzn, stereotypów związanych z płcią czy sposobów utrwalenia w języku myślenia czy doświadczania płci. Zwłaszcza że – jak przekonuje Anna Krawczyk-Tyrpa (1999: 427) – w języku rozróżnienie ze względu na płeć funkcjonuje od zawsze¹⁴.

Wypowiedzi postaci w KorTeDa nierzadko łamią stereotypy związane z płcią, jak na przykład agresywne zachowania mowne wśród młodych kobiet o statusie niskim, czego dobrym przykładem jest fragment wypowiedzi przekupki:

[...] a ja nie wiele myśląc jak złapię za wierzchowisko! zdarłam jej ze łba chustkę!... Dobrze! pani Duckowa mi też pomogła chwyciła za kłaki, a ja dopiero okładam pięścią po pysku. Ale cóż z tego kiedy p. Garusińska to baba jak ciele, więc zamiast chwycić za ręce, stanęła i patrzy (Ku).

W korpusie tego typu wypowiedzi kobiet nie są częste, na przykład wulgaryzmu *cholera* używa Władka w (Asz). Dominują natomiast silnie nacechowane ekspresywnie wypowiedzi zawierające różnego rodzaju interiekcje w języku mężczyzn, na przykład *Idź do diabła!* (Grzegorz, Dh), *A cóż to ty stary psie, podobno się upił?* (Lo-tary, Kmś), *A dajże mi święty pokój!... djabli wiedzą...* (Ignacy, Mk).

Gdy weźmiemy pod uwagę kontekst historyczny, zrozumiałe będą oschłe wypowiedzi matki do córki w sytuacji głodu: *Jak godam, to godam. Coś miała zezryć, toś zezarła, więc kładź się spać*, wypowiedzi tej samej postaci obrazujące silną, zdeterniowaną do działania w obliczu biedy kobietę, która z premedytacją, ale nieświadomie

12 Literatura na temat języka kobiet i mężczyzn, czy szerzej *gender studies*, jest bardzo bogata i zróżnicowana. Zob. więcej w najnowszej książce Jolanty Szpyry-Kozłowskiej (2021).

13 Szczegółowe omówienie zasad anotacyjnych płci zawiera podrozdział 3.2.2.

14 Za: Karwatowska, Jarosz 2012: 60.

zabija swojego syna: *Nie prowadź! Nie razem. Jo sama, i dalej: A kto mi w renke sikire kłod?! Wydorłam ze siebie duse, to mi jej tera nie przywrocaj!, czy rozgoryczonego ojca: Przedom gront, ale krowy nie przedom! Bede krod, ale krowy nie przedom!* (N). Tematyka rozmów, ekspresja werbalna i pozawerbalna (mówienie, spokojne, oschłe, z trudnościami, w sposób wybuchowy itp.), różne reakcje językowe w momencie poznania tragicznej prawdy to tylko wybrane aspekty językowej analizy w kontekście zróżnicowania płci. Powyższe przykłady jedynie w minimalnym stopniu pokazują możliwości historycznych badań językoznawczych zorientowanych socjopragmatycznie, gdy jednym z czynników różnicujących wypowiedzi jest płeć rozmówców.

3.2.2. Zasady oznaczania płci w KorTeDa

Anotacja płci w KorTeDa zakłada przypisanie wypowiedziom pojedynczych postaci jednej z trzech wartości: kobieta, mężczyzna lub nieokreślona płeć postaci według następujących zasad:

- a) **kobieta** – postać płci żeńskiej. Anotowana jest na podstawie nazw własnych, to jest imion: *Aniela, Zuzia, Helena*; form nazwisk odmeżowskich: *Pawłowa, Cześnikowa, Anzelmowa*; nazwisk w formie żeńskiej: *Dulska, Mańkowska* oraz tytułów: *Baronowa, Hrabina*. Na płeć wskazują także nazwy pojawiające się w spisie postaci, w didaskaliach lub tekście głównym, na przykład pokrewieństwa i powinowactwa: *babka, żona, matka, córka, siostra, kuzynka, siostrzenica, pasierbica*; wykonywanych profesji: *guwernantka, krupiarzka, gosposia, kucharka, przekupka, służąca, pokojówka*; identyfikujące płeć żeńską: *kobieta, pani, panna, baba, wdowa, przyjaciółka* itp. W wypowiedziach nadawców płci żeńskiej oraz w wypowiedziach kierowanych do odbiorców płci żeńskiej stosowane są formy przymiotników i czasowników rodzaju żeńskiego lub niemęskoosobowego.
- b) **mężczyzna** – postać płci męskiej. Anotowana jest na podstawie imion: *Karol, Ludwik, Jan, Adolf*; nazwisk w formie męskiej: *Rzetelnicki, Bogacki, Zdzierski*; tytułów: *Baron, Hrabia, Margrabia, Książe*. Płeć sygnalizują ponadto nazwy pokrewieństwa i powinowactwa: *ojciec, mąż, brat, kuzyn, stryj, krewny*; nazwy wykonywanych profesji: *sędzia, porucznik, lekarz, bankier, ekonom, żołnierz, kapelan, lokaj, ogrodnik, masztalerz, fornal*; nazwy identyfikujące płeć męską: *pan, chłop, przyjaciel, właściciel, amant, zalotnik* itp. W wypowiedziach nadawców płci męskiej oraz w wypowiedziach kierowanych do odbiorców płci męskiej stosowane są formy przymiotników i czasowników rodzaju męskiego lub męskoosobowego.
- c) **nieokreślona płeć postaci** – postać, która nie jest określona pod względem płci, na przykład *dziecko, sierota, sługa* (jeśli na podstawie tekstu dramatu nie jest możliwe określenie płci); postaci nierzeczywiste: *zjawa, duch, anioł* itp.

Określenie płci nadawców wypowiedzi zazwyczaj nie było problematyczne. Jedynie trudności dotyczyły przypisania płci wypowiedziom stanowiącym tło dla

głównej partii dialogowej, jak głosy zza ściany czy zza drzwi, pojedyncze okrzyki, głosy modlących się nad zmarłym itp. W takiej sytuacji każdorazowo podejmowano próbę identyfikacji płci nadawców na podstawie możliwie szerokiego kontekstu, na przykład *głos zza drzwi* w dramacie (K) został zinterpretowany jako głos handlarza żydowskiego i wypowiedzi została przypisana wartość: męczyzna. W przypadku nazw postaci, które nie identyfikowały jednoznacznie płci, na przykład *La Flor* (rodzajnik wskazuje na żeńskość, natomiast w dramacie jest to nazwa postaci męskiej [Pc]), *Kukuma* (imię będące neologizmem to nazwa postaci żeńskiej [A]), pomocne okazywały się informacje zawarte w spisie postaci, analiza treści dramatu bądź wypowiedzi danych postaci pod względem wykorzystywanych form gramatycznych. W obecnej wersji KorTeDa nie został wykorzystany znacznik wskazujący na nieokreśloną płeć postaci.

3.3. Status społeczny

3.3.1. Rozumienie kategorii

Status społeczny jest kategorią socjologiczną, wykorzystywaną również w innych naukach społecznych i humanistycznych. Korzysta z niej historia społeczna, potrzebna jest także w badaniach językoznawczych zorientowanych socjolingwistycznie i pragmalingwistycznie, czyli w tych nurtach językoznawstwa, które ściśle wiążą charakterystykę językową użytkowników z kontekstem. Ich historyczne warianty (historyczna socjolingwistyka, historyczna pragmalingwistyka) stają przed szczególnym wyzwaniem, ponieważ dane językowe, które podlegają analizie, muszą być weryfikowane na podstawie ustaleń historycznych co do postrzegania statusu społecznego w danej epoce historycznej, a jednocześnie muszą podlegać niezbędnym uproszczeniom.

Materia społeczna jest niezwykle dynamiczna i wypracowanie opisu, który byłby akceptowalny dla dłuższego okresu (w KorTeDa 150 lat), jest trudne i wymaga szczególnej uwagi. Autorzy współczesnych opracowań socjologicznych (np. Szmatka 1980; Linton 2000) zgodni są co do tego, że wyznacznikami statusu społecznego są poziom wykształcenia, przynależność do grupy zawodowej oraz wskaźniki ekonomiczne. W przypadku materiału historycznego do tego zestawu dołączyć należy dziedziczną przynależność do określonego stanu. Widać tu wyraźne krzyżowanie się parametrów dziedzicznych z tymi, które są wynikiem dokonań jednostki (zob. Ślusarek 2007), przy czym te pierwsze do czasu II Rzeczypospolitej są zdecydowanie istotniejsze. W zasadzie do zniesienia pańszczyzny w 1864 r. niemożliwe było zjawisko tak zwanego awansu społecznego (zob. Leszczyński 2020: 240–329), choć takie próby były podejmowane, co dokumentuje słynna „*Liber chamorum*” Waleriana Nekandy Trepki (1995). Jak pisze Adam Leszczyński, „[a]mbitni plebejusze mieli do

przekroczenia ogromny dystans społeczny, który wyrażał się w głęboko utrwalo-nych kanonach zachowania, stroju i mowy, jeśli już nie w majątku (ale w nim często również)” (2020: 245).

Wiek XVIII i XIX to stopniowy rozpad utrwalo-nych stosunków społecznych, który miał kilka przyczyn. Wśród najważniejszych należy wskazać politykę rządu Stanisława Augusta, publicystykę Sejmu Wielkiego, następnie zmiany wprowadzo- -ne przez zaborców czy wreszcie opinie zachodnioeuropejskich komentatorów poli- -tycznych. Konsekwencje uwłaszczenia chłopów silnie wpłynęły z kolei na przebu- -dowę składu zawodowego stanu najniższego. Jak podaje A. Leszczyński, w 1845 r. na terenie Królestwa było 46 tysięcy robotników, a w 1913 już 400 tysięcy (ibid.: 347). Czy tak duża dynamika miała bezpośrednie przełożenie na zwyczaje językowe? Z pewnością tak, ale był to proces długotrwały, wymagający nie tylko zmian poli- -tycznych i ekonomicznych, ale przede wszystkim mentalnych – te zaś zachodzą zdecydowanie wolniej.

Sygnaly świadczące o wolniejszym przeobrażaniu się języka względem zmian społeczno-politycznych i kulturowych zawierają publikacje Joanny Kuciel-Frydryszak (2022, 2023). Warto przytoczyć kilka przykładowych spostrzeżeń autorki:

Chłopi wciąż mawiają, że pracują „na pańskim”, choć od zniesienia pańszczyzny minęło przeszło sześćdziesiąt lat (2023: 256).

Chociaż obszarnicy ubożeją, przenoszą się do miast, ich wpływy i prestiż pozostają silne, a stosunki z klasami niższymi wciąż oparte na dominacji (2023: 288).

[...] na początku lat pięćdziesiątych [XX wieku – Autorki] [...] [s]łowo „służąca” stało się już słowem obraźliwym. Jednak wciąż trudno było uwierzyć, że epoka służby domowej przeminęła (2022: 155).

W KorTeDa zmiany w języku, jakie dokonały się po zniesieniu pańszczyzny, silnie wybrzmiały w (Cha). Jak powierzchowne to zmiany, świadczą wypowiedzi żony kmiecia – Katarzyny:

Chłop! Chłop!... co było a nie jest, to się nie pise w legestr, pokiśma robili pańszczyzną tośma byli chłopcy, ale teraz jesteśma galant ślachcicami, a wies dobrze ze zaden ślachcic nie może się nazywać Koguciukiem, ale Kogutkiewiczem abo Kogucieńskim (Cha).

A dyć ty osalałeś Marcinku! a kazby się, podzieli kumornicy co nic nie mają, są chłopami i będą chłopami, – a my ze nie potrzebujemy robić, a przeto jesteśma ślacheckiej kondy- -cye. Tak: ja jestem pani, mój jest pan, a ty jesteś pańskie dziecko Marcinku (Cha).

Domaganie się od służby używania formuł *pan*, *pani* czy nowej formy nazwiska nie zmieniają faktu, że wypowiedzi Katarzyny reprezentują język osób o statusie niskim.

3.3.2. Zasady oznaczania statusu społecznego w KorTeDa

Przyjmując założenie o konserwatywności języka jako narzędzia komunikacji społecznej, zastosowano w korpusie trzystopniowy sposób oznaczania statusu społecznego nadawcy wypowiedzi. Jest to, przyjmując pewne uproszczenia, zgodne zarówno z opisem przedstawionym w *Encyklopedii staropolskiej* (Brückner 1939/1990), jak i z opisami współczesnych historyków, którzy zajmują się tym okresem (np. Ichnatowicz, Mączak, Zientara et al. 1999: 429–506; Ślusarek 2007). Koniec XVIII w. to czas, kiedy w podziałach historycznych wyróżnia się cztery stany: szlachtę, duchowieństwo, mieszczaństwo i chłopów. Zdecydowano się przyjąć podział nie odwołujący się do konkretnych stanów, ale wyraźnie pozycjonujący nadawcę wypowiedzi w hierarchii społecznej. Uznano bowiem, że jest on bardziej przydatny w opisie językowym.

Duchowieństwo było jedynym stanem wyróżnianym ze względu na charakter wykształcenia, który mógł mieć różny poziom, natomiast dwa pozostałe czynniki (pochodzenie i status ekonomiczny) ulegały neutralizacji. Stwierdzono na podstawie lektury tekstów korpusu, że w tym przypadku istotne jest rozróżnienie zależne od prestiżu danej funkcji w obrębie tego stanu, to znaczy inny status przypisano plebanowi bądź zakonnikowi (średni), inny zaś biskupowi (wysoki). Również wewnętrzne zróżnicowanie w obszarze stanu szlacheckiego, postępujące wraz z procesami ubożenia szlachty oraz wyodrębniania się ziemiaństwa, a następnie inteligencji, zostało odzwierciedlone poprzez zróżnicowane przypisywanie przedstawicieli tych warstw jednej z trzech wyróżnianych wartości.

W KorTeDa status społeczny postaci określono, biorąc pod uwagę dwa podstawowe kryteria: pochodzenie oraz wykonywany zawód czy codzienne zajęcie. Wiąże się z nimi pośrednio kryterium trzecie, a mianowicie status ekonomiczny. Nie był on jednak w badanym przedziale historycznym decydujący, choć zyskiwał na znaczeniu tym bardziej, im teksty bliższe były współczesności. Status społeczny ustalany był na podstawie spełnienia przez daną postać co najmniej jednego z wymienionych kryteriów. Ważne jest także przyjęte założenie mówiące o dziedziczeniu statusu, to znaczy dzieciom jest przypisywany ten sam status co rodzicom. Poniżej przedstawiono zasady przypisywania poszczególnych wartości statusu społecznego:

- a) **niski** – ze względu na pochodzenie: gospodarz, chłopka, mieszkanka wsi; ze względu na zawód/zajęcie: służący, lokaj, odźwierny, kucharka, pokojowa, służąca, stajenny, koniuszy, parobczak, robotnik itp.
- b) **średni** – ze względu na pochodzenie: mieszczaństwo, ziemiaństwo, zdeklasowana szlachta; ze względu na zawód/zajęcie: radca prawny, prawnik, mecenas, notariusz, właściciel wsi, subiekt, student, kupiec, literat, ekonom, guwernantka, nauczyciel, pleban, porucznik itp.
- c) **wysoki** – ze względu na pochodzenie: arystokracja, magnateria, szlachta; hrabina, hrabia, książę; ze względu na zawód/zajęcie: biskup, oficer, pułkownik, major itp.

d) **nieokreślony** – w przypadku gdy nie jest możliwe ustalenie statusu społecznego postaci na podstawie żadnego z wyżej wymienionych kryteriów.

Źródła informacji o statusie uczestników dialogu stanowią: nazwy postaci wymienione w spisie, na przykład *Wójt*, *Dziedzic* (Jg); didaskalia, na przykład *radca sądu* – wypowiada się jako Pagatowicz (Gr); treść dramatu, na przykład *przedstawicielka koterii*, *zamożna wdowa* (Kw); stosowane w odniesieniu do danej postaci zwroty adresatywne, na przykład *służę wielmożnej Pani* (CM), *WaćPan* (Pc), *moja panienko*, *moja Katarzyno* (Ku); opis postaci przedstawiany przez inną osobę dramatu, na przykład: [...] *prosty szlachcic, gdzieś za piecem chowany; uszy zatykać, gdy zacznie dowcipkować po swojemu...* (Żegocina o bracie swojego męża, PD).

4. Problemy anotacji socjolingwistycznej

Zaprezentowane wyżej zasady znakowania podstawowych wartości socjologicznych zostały wypracowane na podstawie lektury dwóch typów tekstów: tekstów źródłowych tworzących korpus oraz opracowań z zakresu demografii, socjologii kultury i historii społecznej. W trakcie prac nad tym poziomem anotacji pewne kwestie okazały się problematyczne. Miały one naturę zarówno merytoryczną, jak i techniczną.

Zasadnicza trudność w przyporządkowaniu wybranych wartości nadawcom wypowiedzi wynika z historyczności analizowanego materiału. Ze względów oczywistych niemożliwe było zastosowanie ustaleń współczesnej socjologii oraz kompetencji współczesnych użytkowników języka. Wątpliwości dotyczyły przede wszystkim wieku i statusu. Poza czynnikami historycznymi należało wziąć pod uwagę uwarunkowania polityczne, a także te, może rzadziej opisywane w literaturze, które wynikają ze stereotypów, uprzedzeń – indywidualnych i zbiorowych. Dobrym przykładem jest tu dzieło (Ku), w którym głównymi bohaterkami są tytułowe kucharki, często wywodzące się ze wsi, pracujące na służbie w mieście. Zgodnie z wypracowanymi kryteriami przypisano im status niski. Warto jednak wspomnieć, że w ich wypowiedziach pojawia się negatywna ocena ludzi, którzy zostali na wsi, por. Salomea: *Może tam, moja pani i byłby lepszy, ale nie byłby dla mnie honor, żeby za chłopą poszła* (Ku). Indywidualne poczucie wyższej wartości też może przekładać się na sposób wypowiedzania, kreowania postaci. Nieco inny problem pojawia się, kiedy ocena obiektywna (zgodna z wypracowanymi kryteriami) jest niezgodna z oceną wynikającą ze stosunków społecznych w danym czasie. Przykładem tego typu trudności jest postać Żyda karczmarza, występująca w dramacie (N). Zgodnie z kryterium ekonomicznym powinien on być przypisany do grupy o statusie „średni”, jednakże wiedza o pozycji ludności żydowskiej w tamtym czasie (zob. Leszczyński 2020: 313–318) kazała nadać mu status niski.

Jeszcze inny typ trudności, które można by nazwać wewnątrztekstowymi, wynika z nieoczekiwanych zwrotów akcji, które pojawiają się w niektórych dramatach. Przykładem tego typu jest sztuka (Cns), w której Celina, wychowywana jako sierota

przez Starościnę, dowiadyuje się o swoim prawdziwym pochodzeniu: okazuje się, że jest córką ubogiej wyrobnicy. Starościna, wyjawiając tę tajemnicę, ze szczegółami opowiedziała o okolicznościach przyjęcia dziewczyny oraz podjęcia decyzji, by dać jej staranne wychowanie:

Zrazu nie miałam zamiaru wyniesienia cię nad stan twój, ale przywiązanie jakie dla ciebie powzięłam, piękność twoja i niepospolite zdolności, które się wcześniej w tobie rozwijać zaczęły, skłoniły mnie do dania ci tego starannego i wyższego wychowania, które cię w rzędzie zajmujących kobiet stawia (Cns).

W sytuacji gdy znane jest prawdziwe pochodzenie Celiny (stan niski), wiedza o jej wychowaniu i sposobie wypowiedania się stały się podstawą do przypisania jej do postaci o statusie wysokim, jaki ma opiekunka Starościna.

Wykonanie anotacji wiąże się również z koniecznością przygotowania tekstów kompatybilnych z narzędziem korpusowym. Brak konsekwencji autorów dramatów w zapisie nazw postaci powodował, że partie wypowiedane przez jedną osobę rozpoznawane były przez system jako osobne, gdy nie łączyła ich wspólna etykieta, por. Emilia vs. P. Emilia. Didaskalia, które pisane były mniejszą czcionką zamiast w nawiasach, włączane były do kwestii wypowiedanej przez postać, natomiast błędy w numeracji scen powodowały scalanie jednostek tekstu oznaczonych tym samym numerem. Wdrożenia tagerów korpusowych to ostatnie stadium pracy, które poprzedzić musi odpowiednie przygotowanie i skontrolowanie jakości danych. W przypadku dramatów same etykiety nazw postaci, a także oznaczenia aktów i scen muszą zostać oznaczone jako obudowa tekstowa, która nie wlicza się do materiału językowego zdeponowanego w korpusie.

5. Zakończenie

KorTeDa otwiera przed językoznawcami szerokie perspektywy prowadzenia historycznych badań pragmatyngwistycznych. Dotychczas wersja testowa korpusu stała się podstawą do podjęcia badań między innymi nad zróżnicowaniem przerywników leksykalnych w dawnej potocznej komunikacji (Pastuch, Mitrenga, Wąsińska 2024) oraz zorientowanej historycznie funkcjonalnej analizy użycia jednostki *jak Boga Kocham* (Ryś 2023). Materiał korpusowy posłużył także jako materiał porównawczy do badań w zakresie użycia przymiotnika *śladki* w dawnej i współczesnej polszczyźnie (Wąsińska 2023). Kolejne zagadnienia przybliżające historyków języka do udzielenia odpowiedzi na pytanie o to, w jaki sposób mówili dawni Polacy, czekają na opracowanie.

Zastosowana w korpusie anotacja socjolingwistyczna, dzięki której możliwe jest wyekscerpowanie materiału przypisanego do nadawców reprezentujących różny wiek, płeć i status społeczny, pozwala na wykorzystanie historycznych danych

językowych dostosowanych do konkretnie sprofilowanych badań, na przykład w zakresie lingwistyki kulturowej, stylistyki (biolektów, familiolektów, idiolektów), historycznej frazeologii i paremiologii czy *gender studies*. Opracowany korpus zapewne zainteresuje szerokie grono językoznawców i badaczy dyscyplin pokrewnych oraz stanie się impulsem do podjęcia różnorodnych badań z wykorzystaniem zgromadzonego materiału i zastosowanej anotacji.

Źródła

- A: A. Siemoński, *Anarchia domowa czyli moc miłości ojcowskiej*, Kraków 1833.
 Asz: W. Perzyński, *Aszantka. Komedya w 3 aktach*, Lwów 1907.
 Cha: W.L. Anczyc, *Chłopi arystokraci*, Kraków 1851.
 CM: A. Neuwert-Nowaczyński, *Cyrce Mańkowska*, Lwów – Warszawa 1903.
 Cns: F. Skarbek, *Czemuż nie była sierotą!*, Warszawa 1847.
 Dh: A. Fredro, *Damy i huzary*, Lwów 1839.
 Dok: A. Neuwert-Nowaczyński, *Dom kalek*, Lwów – Warszawa 1903.
 Gr: M. Bałucki, *Grube ryby*, Warszawa 1900.
 Jg: M. Janczowicz-Terlecki, *Jesteśmy gotowi*, Lwów 1939.
 K: J.A. Kisielewski, *Karykatury*, Lwów 1903.
 Ka: S. Dobrzański, *Kajcio*, Kraków 1886.
 Kmś: M. Bończa-Tomaszewski, *Kobiety jakich mało na świecie, czyli cnota nadgrodzona*, Wilno 1799.
 Ku: W. Wielogłowski, *Kucharki. Obrazek z obyczajów domownictwa miejskiego*, Kraków 1858.
 Kw: A.K. Czartoryski, *Kawa*, Warszawa 1779.
 Mk: J. Bliziński, *Marcowy kawaler*, Lwów 1879.
 N: K.H. Rostworowski, *Niespodzianka. Prawdziwe zdarzenie w 4-ech aktach*, Kraków 1929.
 Pc: P. Oraczewski, *Polak cudzoziemiec w Warszawie*, Warszawa 1778.
 PD: J. Bliziński, *Pan Damazy*, Złoczów [ca 1925].
 PJ: A. Fredro, *Pan Jowialski*, Warszawa – Kraków 1880.
 Pl: E. Lubowski, *Pogodzeni z losem*, Warszawa 1878.
 Pm: J. Korzeniowski, *Panna mężatka*, Wilno 1845.
 Po: A. Paślowski, *Pocziwski czyli Wybiegi niepospolite*, Wilno 1808.
 R: F. Skarbek, *Roztrzepany*, Warszawa 1847.
 Sr: L.A. Dmuszewski, *Siedem razy jeden*, Wrocław – Kalisz 1823.
 Sz: M. Mann, *Sztuka i miłość*, Poznań 1849.
 Wd: T. Rittner, *W małym domku*, Warszawa 1907.
 Wg: K. Zalewski, *Wycieczka za granicę*, Warszawa 1872.
 Zz: S. Laudyn-Chrzanowska, *Zmarnowane życie*, Lwów 1895.
 Ż: G. Zapolska, *Żabusia*, Warszawa 1903.

Literatura

- ARCHER D., CULPEPER J., 2003, *Sociopragmatic annotation: New directions and possibilities in historical corpus linguistics*, [w:] A. Wilson, P. Rayson, A.M. McEnery (red.), *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*, Frankfurt am Main, s. 37–58.
- ARCHER D., CULPEPER J., DAVIES M., 2008, *Pragmatic annotation*, [w:] M. Kytö, A. Lüdeling (red.), *Corpus Linguistics: An International Handbook*, t. 1, Berlin, s. 613–642.
- BEDNARUK W., 2019, *Dziecko w prawie staropolskim*, „Biuletyn Stowarzyszenia Absolwentów i Przyjaciół Wydziału Prawa Katolickiego Uniwersytetu Lubelskiego” 14, nr 14 (1), s. 7–16. DOI: 10.32084/sawp.2019.14.1-1.
- BRÜCKNER A., 1939/1990, *Encyklopedia staropolska*, t. 1, Warszawa (wyd. fotooffsetowe).
- CULPEPER J., 2009, *An introduction*, „Journal of Historical Pragmatics” 10 (2): *Historical sociopragmatics*, s. 1–8. DOI: 10.1075/bct.31.01cul.
- CULPEPER J., KYTÖ M., 2000, *Data in historical pragmatics: Spoken interaction (re)cast as writing*, „Journal of Historical Pragmatics” 1 (2), s. 175–199. DOI: 10.1075/jhp.1.2.03cul.
- CYBULSKI M., 2003, *Obyczaje językowe dawnych Polaków. Formuły werbalne w dobie średniopolskiej*, Łódź.
- DOBROWOLSKA D., 1992, *Przebieg życia, fazy, wydarzenia*, „Kultura i Społeczeństwo” 36, nr 2, s. 75–87.
- HAJNICH E., MURZYNOWSKI G., WOLIŃSKI M., 2008, *Anotatornia – lingwistyczna baza danych*, [w:] *Materiały V konferencji naukowej InfoBazy 2008, Systemy, aplikacje, usługi*, s. 168–173, Sopot, [on-line:] <http://nlp.ipipan.waw.pl/Bib/haj:mur:wol:o8.pdf>.
- ICHNATOWICZ I., MAĆZAK A., ZIENTARA B. et al., 1999, *Społeczeństwo polskie od X do XX wieku*, wyd. 4, Warszawa.
- JUCKER A.H., 1994, *The feasibility of historical pragmatics*, „Journal of Pragmatics” 22 (5), s. 533–536. DOI: 10.1016/0378-2166(94)90083-3.
- KARWATOWSKA M., JAROSZ B., 2012, *Emulacja płci w tekstach o modzie*, [w:] M. Karwatowska, J. Szpyra-Kozłowska (red.), *Oblicza płci. Język. Kultura. Edukacja*, Lublin, s. 59–72.
- KARWATOWSKA M., SZPYRA-KOZŁOWSKA J. (red.), 2012, *Oblicza płci. Język. Kultura. Edukacja*, Lublin.
- KLEMENSIEWICZ Z., 1961, *Jak charakteryzować język osobniczy?*, [w:] Z. Klemensiewicz, *W kręgu języka literackiego i artystycznego*, Warszawa, s. 204–214.
- KRAWCZYK-TYRPA A., 1999, *Atrybuty męczyzn i kobiet utrwalone w polszczyźnie*, [w:] J. Adamowski, S. Niebrzegowska (red.), *W zwierciadle języka i kultury*, Lublin, s. 427–433.
- KUCIEL-FRYDRYSZAK J., 2022, *Służące do wszystkiego*, Warszawa.
- KUCIEL-FRYDRYSZAK J., 2023, *Chłopki. Opowieść o naszych babkach*, Warszawa.
- KUKŁO C., 2009, *Demografia Rzeczypospolitej przedrozbiorowej*, Warszawa.
- KUKŁO C., ŁUKASIEWICZ J., LESZCZYŃSKA C., 2014, *Historia Polski w liczbach*, t. 3: *Polska w Europie*, Warszawa.
- KYTÖ M., WALKER T., 2006, *Guide to A Corpus of English Dialogues 1560–1760*, Uppsala, „Studia Anglistica Upsaliensia” 130.
- LEECH G., 1997, *Introducing corpus annotation*, [w:] R.G. Garside, G. Leech, A.M. McEnery (red.), *Corpus Annotation: Linguistic Information from Computer Text Corpora*, London – New York, s. 1–18.
- LESZCZYŃSKI A., 2020, *Ludowa historia Polski. Historia wyzysku i oporu. Mitologia panowania*, Warszawa.

- LINTON R., 2000, *Kulturowe podstawy osobowości*, wyd. 2, Warszawa.
- MARCIŃCZUK M., KOCOŃ J., BRODA B., 2012, *Inforex – a web-based tool for text corpus management and semantic annotation*, [w:] N. Calzolari, K. Choukri, Th. Declerck et al. (red.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC' 2012)*, Istanbul, s. 224–230, [on-line:] http://www.lrec-conf.org/proceedings/rec2012/pdf/446_Paper.pdf.
- MITRENGA B., PASTUCH M., WĄSIŃSKA K., 2021, *Możliwości i ograniczenia historycznych badań pragmatycznych*, [w:] M. Mączyński, E. Horyń, E. Zmuda (red.), *W kręgu dawnej polszczyzny*, t. 7, Kraków, s. 163–182.
- PAŁKA P., KWAŚNICKA-JANOWICZ A., 2017, *Przewodnik po elektronicznych zasobach językowych dla polonistów (słowniki, kartoteki, korpusu, kompendia)*, Kraków.
- PASTUCH M., DUDA B., LISZYK K. et al., 2018, *Digital humanities in Poland from the perspective of the historical linguist of the Polish language: Achievements, needs, demands*, „Digital Scholarship in the Humanities” 33 (4), s. 857–873. DOI: 10.1093/llc/fqy008.
- PASTUCH M., MITRENGA B., WĄSIŃSKA K., 2024, *Przerywnik leksykalny w historycznojęzykowych badaniach socjopragmatycznych (na materiale Korpusu dawnych polskich tekstów dramatycznych (1772–1939))*, „Język Polski” CIV, z. 1, s. 93–110. DOI: 10.31286/JP.00488.
- PAWŁOWSKA A., 2014, *Formuły werbalne polskiej etykiety językowej od połowy XVIII do lat sześćdziesiątych XIX wieku. Analiza socjolingwistyczna*, Łódź.
- PĘZIK P., 2013, *Wybrane aspekty reprezentatywności małych i średnich korpusów*, [w:] W. Chlebda (red.), *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*, Opole, s. 45–58.
- REJTER A., 2013, *Płeć – język – kultura*, Katowice.
- RYŚ A., 2023, *Jak frazeologizmy funkcjonują w różnych wariantach i kontekstach? Funkcjonalna analiza materiału historycznojęzykowego na podstawie jednostki jak Boga Kocham*, „LingVaria” XVIII, nr 2 (36), s. 323–337. DOI: 10.12797/LV.18.2023.36.21.
- SLANY K., STRUZIK J., WOJNICKA K. (red.), 2011, *Gender w społeczeństwie polskim*, Kraków.
- SZACKA B., 2011, *Gender i płeć*, [w:] K. Slany, J. Struzik, K. Wojnicka (red.), *Gender w społeczeństwie polskim*, Kraków, s. 19–35.
- SZMATKA J., 1980, *Jednostka i społeczeństwo. O zależności zjawisk indywidualnych od społecznych*, Warszawa.
- SZPYRA-KOZŁOWSKA J., 2021, *Nianiek, ministra i japonki. Eseje o języku i płci*, Kraków.
- ŚLUSAREK K., 2007, *Sytuacja szlachty i chłopstwa na tle przemian struktury społeczeństwa polskiego w XIX wieku*, [w:] P. Franaszek, A.N. Nieczuchrin (red.), *Problemy cywilizacyjnego rozwoju Białorusi, Polski, Rosji i Ukrainy od końca XVIII do XXI wieku*, Kraków, s. 292–298.
- TREPKA W.N., 1995, *Liber generationis plebeorum („Liber chamorum”)*, wyd. 2, Wrocław.
- WALCZAK B., 2017, *Relacja socjolekt–idiolekt na przykładzie języka Zagłoby*, [w:] U. Sokółska (red.), *Socjolekt, idiolekt, idiosyl. Historia i współczesność*, Białystok, s. 411–430.
- WĄSIŃSKA K., 2023, *Przymiotnik słodki w dawnych i współczesnych połączeniach wyrazowych języka polskiego*, „Poradnik Językowy” nr 10, s. 32–47. DOI: 10.33896/PorJ.2023.10.3.
- WOLIŃSKI M., PRZEPIÓRKOWSKI A., 2001, *Projekt anotacji morfosyntaktycznej korpusu języka polskiego*, Warszawa, [on-line:] <http://nlp.ipipan.waw.pl/Bib/wol:prz:01.pdf>.
- ZAŚKO-ZIELIŃSKA M., 2009, *Próba charakterystyki biolektu – język a wiek*, „Studia Linguistica” 28, s. 117–123.

Sociolinguistic Annotation in the Corpus of Old Polish Dramatic Texts (1772–1939)**Abstract**

The article presents the criteria applied in the sociolinguistic annotation in the Corpus of old Polish dramatic texts (1772–1939) (KorTeDa). This level of annotation was about assigning the three basic sociolinguistic variables: age, sex and social status to all the utterances of drama characters (except for collective statements). Six values were selected for determining age (underage, young adult, middle-aged adult, older adult (senior), indeterminate adult, indeterminate); three for sex (female, male, indeterminate); and four for social status (low, medium, high, indeterminate). The article shows the interdependence between those factors and the historical, social and cultural context. Additionally, the authors illustrated substantial and technical problems they faced during the process of this type of annotation. KorTeDa may become an important tool for analysing the history of the Polish language taking into account the methods of corpus linguistics and sociopragmatics.