

# Grzegorz Rudziński

---

## Transkrypcja tekstów pisanych przez cudzoziemców na potrzeby korpusu uczniowskiego

---

Acta Universitatis Lodzensis. Kształcenie Polonistyczne Cudzoziemców 18,  
379-385

---

2011

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

*Grzegorz Rudziński*  
Uniwersytet Łódzki

### **TRANSKRYPCJA TEKSTÓW PISANYCH PRZEZ CUDZOZIEMCÓW NA POTRZEBY KORPUSU UCZNIOWSKIEGO**

Słowa kluczowe: jpjo, korpus językowy, pisownia

Studenci-cudzoziemcy używają polszczyzny w specyficzny sposób. SJPdCUL zgromadziło zasób tekstów pisanych, które mogą być umieszczone w korpusie językowym. Jego przygotowanie właśnie się zaczęło. Rękopisy cudzoziemców mają jednak pewną graficzną specyfikę, a niektóre błędy zapisu prowadzą do efektów humorystycznych. Trudności wynikają często z nieodróżniania specyficznie polskich liter „ą” i „ę”; mylenia polskiego „l” z łacińskim „t” oraz brania polskiego „z” za „z”. Wszystkie te zjawiska powinny być dostępne dla badaczy używających korpusu.

Czynniki konstytuujące polszczyznę studentów-cudzoziemców (psc)<sup>1</sup> jako odmianę współczesnego języka polskiego mają charakter społeczny, językowy, kulturowy, psychologiczny, dydaktyczny i administracyjny (por. Rudziński 2010). Zbadanie psc jest możliwe na drodze bieżąco prowadzonych obserwacji lub dzięki wykorzystaniu archiwalnych zapisów zgromadzonych w ośrodkach kształcenia cudzoziemców. Dla uzyskania pełnego obrazu psc konieczne jest postępowanie obiema ścieżkami, a każda z nich swą specyfiką będzie kształtować metodę badań.

Instytucją mającą najstarsze w Polsce, nieprzerwane tradycje w zakresie przygotowania cudzoziemców do podjęcia studiów w krajowych uczelniach jest Studium Języka Polskiego dla Cudzoziemców Uniwersytetu Łódzkiego<sup>2</sup> (por. Rakowski 2010; Wielkiewicz-Jałmużna 2008). Szczerpłość miejsca zajmowanego przez tę jednostkę, po przeniesieniu do obecnej siedziby, nie pozwoliła jednak na zachowanie wszystkich gromadzonych od 1952 r. materiałów. Z archiwalnych prac egzaminacyjnych pisanych przez cudzoziemców zachowały się ich

---

<sup>1</sup> Proponuję posługiwanie się skrótowcem psc (pisany bez kropki zarówno w mianowniku jak i w przypadkach zależnych) w znaczeniu ‘polszczyzna studentów-cudzoziemców’. W środowisku polonistów-glottodydaktyków na podobnej zasadzie funkcjonuje od lat skrótowiec jpjo ‘język polski jako obcy’.

<sup>2</sup> Dalej w tekście SJPdCUL.

komplety z lat 2004–2009. Gromadzenie kolejnych zbiorów prac egzaminacyjnych nie skutkuje już na szczęście usuwaniem najstarszych z nich, ponieważ zostały objęte procedurami badawczymi. Prace te stanowią cenny materiał ukazujący psc w fazie przechodzenia cudzoziemców z poziomu A2 na B1 i w fazie przechodzenia z B1 na B2, a w odniesieniu do studentów o polskim lub słowiańskim pochodzeniu także poziomów wyższych.

SJPDcUŁ gromadzi w swoim archiwum cztery grupy dokumentów, które mogą być źródłem wiedzy nt. psc:

- podania<sup>3</sup> (z dodatkiem niewielkiej liczby innych pism użytkowych) kierowane w różnych sprawach do dyrekcji Studium lub za jej pośrednictwem do innych władz zwierzchnich;
- arkusze ocen pozwalające stwierdzić, jakie były opinie nauczycieli kształcących cudzoziemca podczas pobytu na kursie na temat jego rozwoju językowego i przygotowania merytorycznego;
- semestralne prace egzaminacyjne z języka polskiego ogólnego;
- końcowe prace egzaminacyjne w zakresie jednej z czterech podstawowych specjalizacji<sup>4</sup>: ekonomicznej, humanistycznej, medycznej lub technicznej.

Dwie ostatnie grupy dokumentów to teksty, które nawet w pobieżnej ocenie dają duże możliwości uzyskania wiedzy nt. psc.

Polszczyzna studentów-cudzoziemców jest z pewnością zróżnicowana wewnętrznie wg podziałów posługującej się nią grupy osób, lecz wstępne obserwacje (Rudziński 2010: 112) wskazują także na istnienie cech wspólnych, łączących osoby z odległych sobie kulturowo i językowo obszarów. Cudzoziemcy przyjeżdżający do Łodzi, by podjąć naukę na kursie przygotowującym ich do studiów w polskich szkołach wyższych, składają się z kilku grup różniących się od siebie cechami określającymi stopień znajomości lub stopień dostępności polszczyzny. Nie chodzi tu przy tym o podziały sprowadzające się do zróżnicowania poziomu sprawności w języku, tradycyjnie opisywane terminami: początkujący, średni, zaawansowany. Sieć podziałów jest znacznie gęstsza i wyznaczają ją takie czynniki, jak: dotychczasowe doświadczenia szkolne, idące zwykle w parze z przynależnością do kręgu kulturowego, ogólny poziom wykształcenia, stopień znajomości polskich (lub choćby europejskich) realiów, typ odebranego wychowania (niekiedy jego brak), rodzaj i poziom motywacji, liczba języków opanowanych wcześniej w trybie nauki szkolnej (co sprowadza się do dysponowania – lub nie – własnymi efektywnymi sposobami uczenia się języka), wresz-

<sup>3</sup> W pisaniu podań i innych pism użytkowych cudzoziemcy często korzystają z pomocy osób trzecich, zatem ta grupa tekstów tylko pod pewnymi warunkami może być wykorzystana do orzekania o polszczyźnie podpisujących się pod nimi osób.

<sup>4</sup> Możliwe są pewne warianty w obrębie każdej z nich, np. egzamin konkretnej grupy humanistycznej może być silniej ukierunkowany na zagadnienia związane z prawem, psychologią, pedagogiką czy kulturą – w zależności od profilu przyszłych studiów, jaki reprezentuje zdająca grupa. Podobne zabiegi mogą być prowadzone we wszystkich specjalizacjach.

cie przynależność do rodziny językowej. Ten ostatni czynnik, który mógłby decydować o sukcesie/porażce w przyswajaniu polszczyzny, w świetle nauczycielskich doświadczeń nie zawsze jest w skali kursu trwającego 9 miesięcy czynnikiem najważniejszym. Niedostateczna motywacja częściej jest przyczyną porażki niż duży dystans między językiem rodzimym studenta a polszczyzną; natomiast bliskość języka studenta i polszczyzny nie gwarantuje sukcesu (por. Skalska, Skalski 1995; Rudziński 1995). Tak duża liczba zmiennych powoduje, że w materiale językowym uzyskanym z prac egzaminacyjnych możemy oczekiwać dużego zróżnicowania zjawisk. Należy więc przyjąć metodę, która nad tym bogatym zróżnicowaniem pomoże nam zapanować. Jacek Waliński (2005: 40), powołując się na Vincenta Ooi (1998), pisze: „Niektórzy językoznawcy zakładają, że występują trzy podstawowe sposoby dokumentowania i uzasadniania danych językowych: 1) introspekcja oparta na intuicji językowej; 2) odnotowanie cytatu; 3) sprawdzenie danych korpusowych”. Oczywiście pierwsza z wymienionych metod w przypadku badacza Polaka w odniesieniu do psc w ogóle nie wchodzi w rachubę. O drugiej z nich J. Waliński pisze, że przy jej zastosowaniu powstają pytania: „jak przedstawiało się odnotowane użycie w szerszym kontekście oraz na ile zauważony przykład jest naturalny dla danej społeczności językowej – czy nie jest użyciem jednorazowym wymyślonym pod wpływem jakiegoś impulsu czy fantazji?” (Waliński 2005). Podzielając te wątpliwości, dodam, że w psc odnotowanym zjawiskiem może być także błąd językowy wynikający z trudnej niekiedy do wyłowienia przyczyny – upatrywanej wtedy zwykle w interferencji zewnątrzjęzykowej. Pozostaje nam zatem użycie metody korpusowej, która zdaje się być najlepszym rozwiązaniem nie tylko z braku innych. Metoda ta pozwala obserwować obszerny materiał językowy i dawać solidną ilościową podstawę do wyciągania wniosków, przy równoczesnym utrwaleniu wszystkich zmiennych cech badanego materiału, co jest szczególnie cenną cechą korpusów uczniowskich – czyli odnotowujących zjawiska występujące w wypowiedziach i tekstach osób uczących się języka. Możliwą procedurę przepisywania prac egzaminacyjnych jednego rocznika wzięła na siebie pani Eliza Rogut (studentka specjalizacji glottodydaktycznej, prowadzonej przez Katedrę Lingwistyki Stosowanej i Kulturowej UŁ), która przygotowuje na podstawie tego materiału pracę magisterską pod opieką prof. dr hab. Marii Witkowskiej-Gutkowskiej. Pierwszy krok w kierunku stworzenia korpusu uczniowskiego ze zgromadzonych w SJPdCUŁ tekstów mamy już więc za sobą.

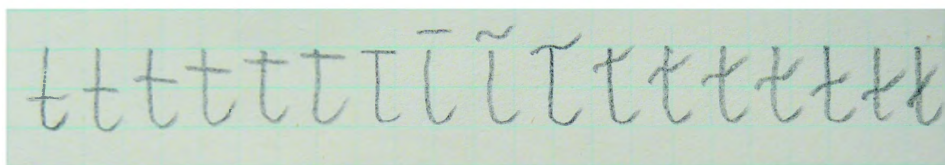
Jedną z cech analizowanego materiału – wyraźnie widoczną już w pierwszym kontakcie – jest występująca w niektórych pracach egzaminacyjnych (ujawniająca się w różnym stopniu) nieporadność zapisu. Skutkuje ona trudnościami z odwzorowaniem oryginalnej grafii autora pracy w wersji odpowiadającej jakiemuś edytorowi tekstu, który może zostać użyty do zbudowania bazy danych. Nie chodzi tu przy tym o typowe błędy ortograficzne, bo te można zapi-

sać bez trudu, usunąwszy z aktywnego menu automatyczną korektę. Staranność obserwacji wystarczy też do odnotowania braku lub nadmiaru znaków diakrytycznych, nad których polskim zestawem niektórzy cudzoziemcy nie panują. Kłopotem jest wynikająca z indywidualnego sposobu pisania niejednoznaczność zapisów. Widzimy ją, gdy nie możemy zinterpretować poszczególnych liter zapisu, a kontekst nie podpowiada intencji autora. Z podobnym stanem mają do czynienia badacze dawnych rękopisów – najnowszy przykład to nowe odczytanie fragmentu *Legendy o św. Aleksym*, gdzie Zofia Wanicowa (2009: 260) wykazuje, że 182. werset *Legendy*: „každy nań pomyje, złą wodę lał” to raczej „každy nań pomyje złą wola 1a1” czyli ‘kierując się złymi intencjami lał’, a zmiana rozumienia wynika z przyjrzenia się katalogowi średniowiecznych abrewiatur, w którym autorka wskazuje możliwość mylenia liter „d” i „l” przy niestarannym zapisie<sup>5</sup>. Nie chodzi tu przy tym znowu tylko o fatalny charakter pisma. Niektóre litery używane w Polsce są na tyle specyficzne, że cudzoziemcy muszą nauczyć się ich pisania, nawet jeżeli wcześniej znali alfabet łaciński – co w przypadku słuchaczy SJPdCUŁ nie zawsze ma miejsce. Przykładowo: wielu cudzoziemców ma zwyczaj zapisywania litery „z” z poziomą kreską, jako „z̄”, co Polak identyfikuje z literą „ż”. Do identyfikacji zapisu potrzebne jest tu oswojenie się z odczytywanym charakterem pisma. Jak jednak odczytać zapis: „Na wieś człowiek chodzi na droga, jaki on chce, ale w miasto on musi chodzi na zebry”. Oczywiście temat pracy (*Wolisz miasto czy wieś?*) i zasada przychylności wobec ucznia nakazują nam rozumieć to jako ‘Na wsi człowiek chodzi drogą jak chce, ale w mieście musi korzystać z przejść dla pieszych’ – bo mniej prawdopodobna w przypadku uczącego się w warunkach akademickich cudzoziemca jest znajomość zwrotu „chodzić na zebry” niż rzeczownika „zebra” w przytoczonym znaczeniu. Inną przyczynę nieczytelności znajdziemy w wielu tekstach pisanych przez cudzoziemców w odniesieniu do liter „ą” i „ę”. Niektórzy studenci zdają sobie sprawę z ważności odróżniania tych liter, ale nie panując nad fleksją (co mogłoby się ujawnić w użyciu końcówki fleksyjnej -ą lub -ę), piszą oba znaki tak, by sprawdzający nauczyciel mógł interpretować je według własnej (najczęściej dobrej) woli. Prawdopodobne jest także, że obserwowane zapisywanie głoski [ɔ] za pomocą litery „o” z ogonkiem (tak jak w starszej wersji zapisu fonetycznego, co czyni ją bardziej podobną do „ę”), wynika z dobrze odbieranego brzmienia tej głoski, która cudzoziemcom nie kojarzy się z użyciem znaku „a”.

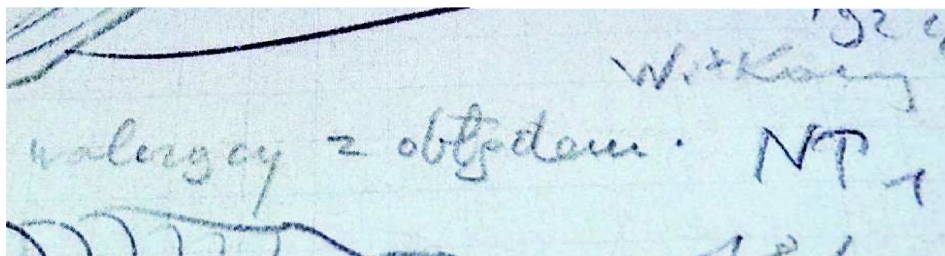
Największe jednak trudności w zakresie specyficznie polskiej pisowni mają cudzoziemcy z literą „l”. W wielu pracach nie można jej odróżnić od litery „t”. Poniższa ilustracja przedstawia niektóre sposoby pisania tych dwóch liter, przy

<sup>5</sup> Ze zjawiskiem tym większość Polaków spotyka się podczas prób odcyfrowania treści niechlujnie napisanych recept, gdy nieświadomi konsekwencji swojego działania lekarze narażają tym swych pacjentów na utratę zdrowia – o ile nie życia.

czym nie istnieje reguła, że zapisy z lewej strony to zawsze „t”, a zapisy z prawej to zawsze „l”.



Zagadnienie to jest szczególnie wyraźne w pracach studentów arabskich, którzy często słabo władając alfabetem łacińskim, muszą przedstawiać sobie kierunek zapisu, co zaostrza kłopot z utrzymaniem jego staranności. Dochodzi do tego powszechna u tych Arabów, których uczyłem, wynikająca ze specyfiki ich alfabetu, skłonność do nieodróżniania liter wielkich i małych, z możliwością użycia tych pierwszych także w środku wyrazu; a wtedy majuskuła „T” może niczym nie różnić się od minuskuły „t”, jeżeli ktoś pisze „l”, „l” i „t” nie jako odwróconą laseczkę, ale jako pionową kreskę. Wieloletnie obcowanie z tekstami pisanyymi przez cudzoziemców buduje we mnie przekonanie, że tak jak indywidualnym realizacjom głosek odpowiada w systemie języka jeden idealny fonem, tak w tekstach pisanych przez członków jednej wspólnoty graficznej pojawiać się może osadzona w umysłach piszących i czytających idea literemu. Pozwalam sobie rozważać istnienie wspólnot graficznych i literemów na podstawie obserwacji, z których wynika, że np. cechy pisma wszystkich znanych mi frankofonów (z kilku krajów) wykazują pewne regularne podobieństwa w obrębie tej grupy piszących (np. sposób pisania „z”), a cechy pisma anglofonów (także z kilku krajów) wykazują inne specyficzne cechy (np. sposób pisania „r”). Podczas transkrypcji tekstów z rękopisów do edytora powstają wątpliwości, które są najsilniejsze, gdy zapis nie realizuje w istocie żadnego z dwóch opozycyjnych w polskiej wspólnocie graficznej literemów. Być może stosując zasadę przychylności można uznać, że jeśli nawet osoby od dziecka posługujące się polską pisownią, jak np. Witkacy, którego próbkę pisma przedstawia poniższa ilustracja (za Zgodzińska 2006) piszą tak, że nietrudno w ich napisach pomylić „a” z „ę” i „l” z „t”, to ocalenie w transkrypcji specyfiki grafii cudzoziemców nie jest celowe.



Zauważmy jednak, że konsekwencją porzucenia tego zamiaru już na etapie gromadzenia materiału w korpusie może być przeoczenie tych trudności w nauce języka polskiego, które właśnie z nieodróżniania liter wynikają, a które charakteryzują psc.

Zauważmy, że w obrębie samego słownika minimum polszczyzny w wersji Zofii Kurzowej i Haliny Zgólkowej (1992) – a więc w warstwie leksyki opanowanej wcześniej, gdy niektórzy uczący się jeszcze nie do końca swobodnie piszą (przypomina to czasem mozolne rysowanie liter), z nieodróżniania „t” od „l” wynikać mogą nieporozumienia lub niezamierzone efekty komiczne przy takich parach wyrazów, jak: siła : sita; \*plaka : ptaka; mały : maty; umyły : umyty; \*małka : matka; data : dała; brat : brał; gotowa : \*gołowa czy piłka : pitka.

Nie ma znaczenia, że niektóre człony tych par nie istnieją w języku polskim. Nieprecyzyjnie zapisane w notatkach studenta utrwalają się. Znany jest mi przypadek, gdy pilny student po dniu absencji w klasie nauczył się deklinacji czasu przeszłego na podstawie notatek (nołalek/notalek/nołatek?) kolegi w wersji: „ja czytałem; ty czytateś; on czytat”. Nauczyciel jpjo jest przyzwyczajony, że formy obce polszczyźnie pojawiają się w tekstach jego studentów. Zdarza się też, że kłopotliwemu zapisowi towarzyszą inne błędy i przykładowo: błędnie utworzona forma od „chcieć”: „ona chcęła” zostaje zapisana jako „hcęta”, a czytający młody nauczyciel polonista zastanawia się w pierwszym odruchu, skąd cudzoziemiec znalazł formę dialektalną i dlaczego połączył zaimek 3. os. l. p. z formą 2. os. l. m. Jeśli więc mamy zachować specyfikę tego rodzaju zapisów w korpusie uczniowskim, musimy pomyśleć o odpowiednim znakowaniu ułomnie realizowanych liter. Technicznie nie jest to trudne, wystarczy użycie innej barwy lub innego kroju czcionki, warto jednak zbudować korpus tak od strony informatycznej, by z miejsc wskazujących problematyczne interpretacyjnie zapisy istniało odesłanie do fotografii oryginalnego zapisu. Wtedy każdy językoznawca badający psc poprzez korpus będzie miał możliwość poznania także graficznej warstwy charakterystycznych cech tej odmiany polszczyzny.

#### BIBLIOGRAFIA

- Kurzowa Z., Zgólkowa H., 1992, *Słownik minimum języka polskiego*, Poznań.
- Ooi V. B. Y., 1998, *Computer Corpus Lexicography*, Edynburg.
- Rakowski B., 2010, *Studium Języka Polskiego dla Cudzoziemców Uniwersytetu Łódzkiego (1958-2008)*, [w:] *Teksty i podteksty w nauczaniu języka polskiego jako obcego – 2. Jubileusz 50-lecia Studium Języka Polskiego dla Cudzoziemców UŁ*, „Acta Universitatis Lodziensis. Kształcenie Polonistyczne Cudzoziemców”, z. 17, red. G. Zarzycka, G. Rudziński, Łódź, s. 9–17.
- Rudziński G., 2010, *Stan i perspektywy badań polszczyzny studentów-cudzoziemców (psc)*, [w:] *Współczesna polszczyzna w badaniach językoznawczych. Od gramatyki do języka w komunikacji*, red. P. Zbróg, Kielce, s. 107–113.

- Rudziński G., 1995, *Wysoka sprawność komunikacyjna jako problem kształcenia językowego*, [w:] *Kształcenie sprawności komunikacyjnej Polaków ze Wschodu*, red. J. Mazur, Lublin, s. 59–66.
- Skalska A., Skalski T., 1995, *Pułapka komunikatywności*, [w:] *Kształcenie sprawności komunikacyjnej Polaków ze Wschodu*, red. J. Mazur, Lublin, s. 49–57.
- Waliński J., 2005, *Typologia korpusów oraz warsztat informatyczny lingwistyki korpusowej*, [w:] *Podstawy językoznawstwa korpusowego*, red. B. Lewandowska-Tomaszczyk, Łódź, s. 27–41.
- Wanicowa Z., 2009, *Warsztat badawczy leksykografa historyka*. „Język Polski”, nr 4–5, s. 255–262.
- Wielkiewicz-Jałmużna D., 2008, *Studium Języka Polskiego dla Cudzoziemców Uniwersytetu Łódzkiego (1952–2002)*, „Acta Universitatis Lodziensis. Kształcenie Polonistyczne Cudzoziemców”, z. 15, Łódź.
- Zgodzińska B., 2006, *Witkacy. Dziwność istnienia*, Marki.

### Summary

Foreign students studying in Poland use the Polish language in a specific way. The School of Polish for Foreigners at the University of Lodz has collected some written texts that can be used in a language corpus. The preparation of such a corpus has just started. Handwritings, however, have their visual characteristics. Texts written by foreign students include mistakes that lead to humorous effects. Frequent difficulties come from unrecognized differences between specifically Polish letters “ą” and “ę”; between Polish letter “ł” and Latin “t” or Polish “ź” pronounced as [zh] in contrast to “z” pronounced as [z]. All these visual effects should be available for researchers using the corpus.