

KAROL OPARA

Rymy częstochowskie w poezji polskiej — ujęcie ilościowe¹

1. Wprowadzenie

Odbiór poezji w znacznym stopniu zależy od jej właściwości brzmieniowych, takich jak: akcenty, intonacja czy też rymy i rytm. Ich analiza pod względem językowym przeprowadzana jest od tysięcy lat. Ilościowa ocena struktury wiersza, zwłaszcza zaś utworów o znacznych rozmiarach, stała się znacznie efektywniejsza po opracowaniu narzędzi do komputerowego przetwarzania tekstu.

Cechy brzmieniowe są łatwiejsze w automatycznej analizie niż semantyka, której wyrafinowanie i wieloznaczność jest nieraz wyzwaniem, nawet dla specjalistów. Ich statystyczne opracowanie, wzbogacone o analizę składniową, stanowi część nowego działy badań zwanego komputerowym wspomaganiami poezji (ang. CAP — Computer-Aided Poetry). Wśród licznych zastosowań tego typu podejścia wymienić należy analizę, ocenę, tworzenie, przekłady na inne języki oraz parafrazowanie poezji.

Jednym z kluczowych wyzwań stojących przed CAP-em jest opracowanie obiektywnych miar jakości poezji, a ściślej warsztatu poetyckiego. Próby różnego rodzaju oceny poezji i jej twórców były podejmowane wielokrotnie, prowadząc niekiedy nawet do konfliktów personalnych między poetami, np. Mickiewicza ze Słowackim. Problem ten pojawia się w dyskusji między nauczycielem i uczniem w *Ferdydurke* (Gombrowicz, 1937):

— A zatem dlaczego Słowacki wzbudza w nas zachwyty i miłość? (...) Dlatego, panowie, że Słowacki wielkim poetą był! (...)

¹ Badanie zrealizowano dzięki stypendium naukowemu w ramach projektu pn. „Technologie informacyjne: badania i ich interdyscyplinarne zastosowania” realizowanego ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego, Program Operacyjny Kapitał Ludzki (Umowa nr UDA-POKL.04.01.01-00-051/10-00).

- Boże, ratuj, jak to mnie zachwyca, kiedy mnie nie zachwyca?
- Jak to nie zachwyca Gałkiewicza, jeśli tysiąc razy tłumaczyłem Gałkiewiczowi, że go zachwyca.

Czy Słowacki w istocie wielkim poetą był? Aby ułatwić odpowiedź na to pytanie, zbadano jakość i finezję jego rymów, świadcząca o poziomie jego warsztatu poetyckiego. Język polski jest silnie fleksyjny, co oznacza, że części mowy w tej samej odmianie mają identyczne końcówki. Tworzą zatem trywialne, gramatyczne rymy, w języku potocznym zwane rymami częstochowskimi².

Wzrost zainteresowania środowiska naukowego metodami CAP-u widać na przykładzie recenzentów przesyłających swoje uwagi wierszem (Genzel i in. 2010 oraz Anonimowy recenzent 2010). Mimo to słabo zbadanym obszarem wciąż pozostaje poezja w językach fleksyjnych, w których rymy zależą od odmiany. Celem niniejszej pracy jest próba uzupełnienia tego braku dla języka polskiego. W części 2 wprowadzono ilościową miarę udziału rymów częstochowskich w wierszu (bądź zbiorze wierszy) nazwaną punktami częstochowszczyzny, albo też Cz-punktami. W części 3 przeprowadzono analizę porównawczą współczesnej oraz dawnej poezji polskiej. Po krótkim podsumowaniu pracy zamieszczono dwa załączniki, z których pierwszy opisuje metodę wykrywania rymów, a drugi szczegóły procedury testowania statystycznego.

2. Cz-punkty i poemat referencyjny

Pierwszym krokiem badania było wykrycie rymów i przeprowadzenie ich analizy morfologicznej, co szczegółowo opisano w załączniku A. Otrzymane w ten sposób pary rymujących się słów oraz listy możliwych znaczników morfologicznych przedstawiono w tabeli 1. Zaczerpnięte one zostały z inwokacji do *Pana Tadeusza* (Mickiewicz 1834):

Panno Świąta, co Jasnej bronisz Częstochowy
I w Ostrej świecisz Bramie! Ty, co gród zamkowy
Nowogródzki ochraniasz z jego wiernym ludem!
Jak mnie dziecko do zdrowia powróciłaś cudem (...)

Znaczniki przedstawione w tabeli 1 opisują dopuszczalne przez polską gramatykę formy morfologiczne wyrazu rozpoznane za pomocą analizatora *Morfeusz* (Saloni i in. 2011). Na przykład *Częstochowy* może być rzeczownikiem (subst) rodzaju żeńskiego (f) w dopełniaczu (gen) liczby pojedynczej (sg), np. *bronieć Częstochowy*. Zapisuje się to

² Pojęcie rymów częstochowskich wywodzi się od pobożnych książek dla pielgrzymów z poezją religijną wątpliwej jakości (Piersiak 2008). Czy wena poetycka w istocie zależy od bliskości ikony Czarnej Madonny? Jej liczne cuda zostały uznane aktem papieża Klemensa IX już w 1716 roku. Niestety, weryfikacja zależności warsztatu poetyckiego od odległości od ikony wydaje się być trudna, gdyż współcześni poeci częstochowscy unikają pisania rymowanych wierszy. Widać to np. w magazynie „Galeria” wydawanym przez częstochowskie Literackie Towarzystwo Wzajemnej Adoracji LiTWA (Galeria 2012).

w skrócie jako subst:sg:gen:f. Słowo to może być również formą mianownika, biernika lub wołacza (nom.acc.voc) liczby mnogiej (pl), np. *widzę trzy Częstochowy*, co zapisuje się jako subst:pl:nom.acc.voc:f. Z kolei wyraz „zamkowy” jest przymiotnikiem (adj) w liczbie pojedynczej (sg), mianowniku, bierniku lub wołaczu (nom.acc.voc) w rodzaju męskim osobowym (m1), męskim zwierzęcym (m2) lub męskim rzeczowym (m3) i stopniu równym (pos). Szczegółowe omówienie znaczników morfologicznych zawarto w pracy (Woliński, 2003).

Tabela 1. Analiza morfologiczna rymów z inwokacji do *Pana Tadeusza* Adama Mickiewicza z przypisanymi punktami częstochowszczyzny; prawidłowe znaczniki wyróżniono tłustym drukiem

Słowo 1	Znacznik 1	Słowo 2	Znacznik 2	Cz-punkty	Rym częst.
Częstochowy	subst:sg:gen:f subst:pl:nom.acc.voc:f	zamkowy	adj:sg:nom.voc:m1. m2.m3:pos adj:sg:acc:m3:pos	0	nie
ludem	subst:sg:inst:m3	cudem	subst:sg:inst:m3 subst:sg:inst:n2	50	tak
opiekę	subst:sg:acc:f fin:sg:pri:perf	powiekę	subst:sg:acc:f	50	tak
progu	subst:sg:gen:m3 subst:sg:loc:m3 ubst:sg:voc:m3	Bogu	subst:sg:dat:m1 subst:sg:loc:m1	0	nie
biała	subst:sg:nom.voc:f adj:sg:nom:f:pos adj:sg:voc:f:pos	pała	subst:sg:gen:m3 subst:sg:nom:f fin:sg:ter:imperf	50	nie
ugoru	subst:sg:gen:m3	dworu	subst:sg:gen:m3	100	tak

Aby ocenić częstość rymów gramatycznych w wierszu, wprowadzono miarę zwaną punktami częstochowszczyzny lub Cz-punktami. W przypadku gdy oba rymujące się słowa mają tylko jeden, wspólny znacznik morfologiczny, z pewnością tworzą one rym częstochowski i przypisywana jest im wartość Cz = 100 punktów. Jeżeli część wspólna obu list znaczników jest pusta, przyznawane jest zero punktów częstochowszczyzny Cz = 0. W pozostałych przypadkach trudno określić, czy rym jest gramatyczny. Przykładowo, może się tak zdarzyć, że wyrazy „opiekę” i „powiekę” są rzeczownikami rodzaju żeńskiego w bierniku, np. *pod Twoją opiekę (...) podniosłem powiekę*.

Z drugiej strony „opiekę” może być czasownikiem w formie nieprzeszłej, np. *opiekę mięso na rożnie*. Rozstrzygnięcie, czy rym ten jest gramatyczny, wymaga zatem wybrania prawidłowych znaczników. Proces ten, nazywany ujednoznacznianiem, oparty jest o analizę składniową. Jej automatyczne przeprowadzenie dla poezji jest trudne i wykra-

cza poza zakres tej pracy. W związku z tym, w przypadku gdy listy znaczników dla rymu zawierają zarówno elementy wspólne, jak i różne, przypisywana mu jest kompromisowa wartość $Cz = 50$ punktów. Przykład tego podejścia zobrazowano w tabeli 1, w której prawidłowe znaczniki zostały ręcznie wyróżnione wytłuszczonym drukiem. Zaznaczono również odpowiadającą im liczbę Cz -punktów oraz oznaczono, czy oryginalne rymy w istocie są gramatyczne.

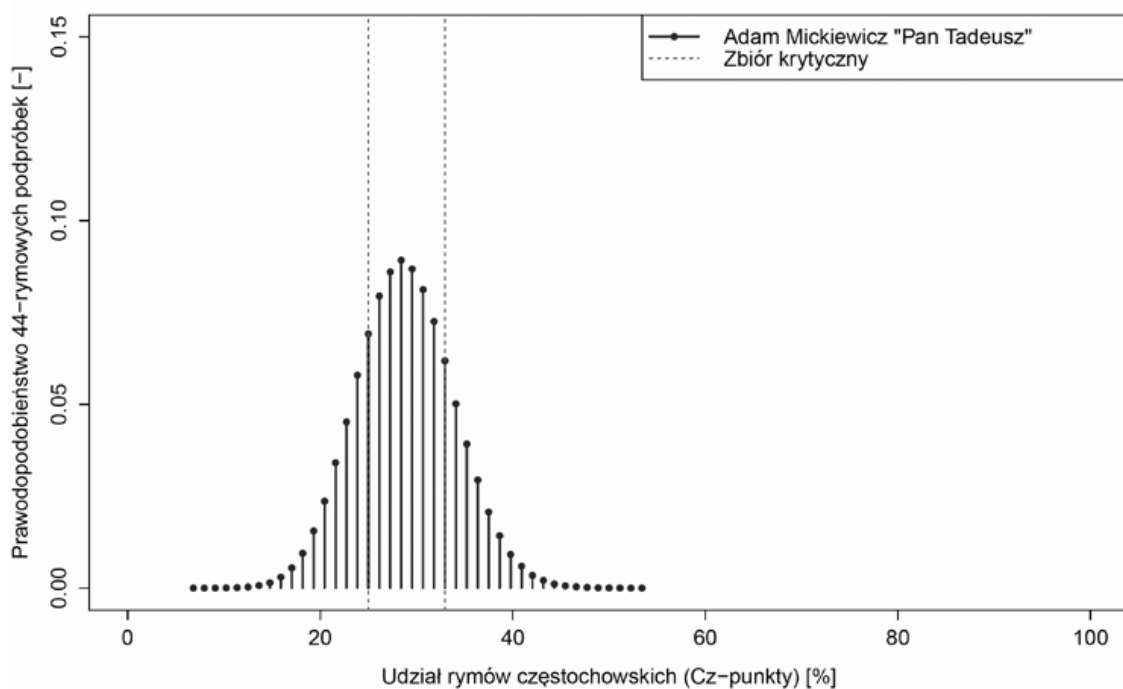
Ile częstochowszczyzny powinno być w wierszu? Zgodnie z przyjętym kanonem nadużywanie rymów gramatycznych obniża wartość dzieła. Z drugiej strony, kunsztowną strofę, pozbawioną rymów częstochowskich, spotyka się jedynie w krótkich wierszach. W poematach jest to praktycznie niemożliwe, gdyż wywoływałoby to efekt sztuczności językowej. Oznaczałoby to również rezygnację z powtórzeń jako środka wyrazu. Nie chodzi zatem o całkowite uniknięcie rymów częstochowskich, tylko o to, by nie były one dominujące w dziełach wierszowanych. Pojawia się więc pytanie, jaki udział rymów gramatycznych uznać należy za konsekwencję właściwości języka, nie zaś nieporadności warsztatowej twórcy? Wybór takiej „dopuszczalnej” ilości ma charakter nieco arbitralny. Zdecydowano się oprzeć go na dziele referencyjnym, w którym względna ilość rymów gramatycznych jest na poziomie charakterystycznym dla poezji wysokiej próby. Do pełnienia tej funkcji wybrano poemat Adama Mickiewicza „Pan Tadeusz”, który uznawany jest za epopieję narodową. Ponadto stanowi jedno z najdłuższych dzieł wierszowanych w literaturze polskiej (dwanaście ksiąg i „Epilog” to łącznie 9850 wersów), co pozwala w wiarygodny sposób przeprowadzić analizy statystyczne.

Z analizy utworu referencyjnego wynika, że średni udział rymów gramatycznych (częstochowskich) wynosi 29 procent. Tak więc za referencyjną wartość częstochowszczyzny uznano $Cz_{ref} = 29$.

Porównywanie ilości rymów gramatycznych w różnych dziełach z poematem referencyjnym wymaga opracowania testu statystycznego, który pokaże, czy ich częstość jest istotnie różna od wartości referencyjnej Cz_{ref} . Z przyczyn obliczeniowych rzetelne porównanie wymaga wzięcia pod uwagę różnicy w długości wierszy. Z tego względu test statystyczny zdecydowano się oprzeć na podpróbkach z tekstu „Pana Tadeusza” o ustalonej długości czterdziestu czterech rymów.

Ta, nieco arbitralnie przyjęta, stała odzwierciedla wybór Mickiewicza przedstawiony w „Wielkiej Improwizacji” (Mickiewicz 1832):

Wybawca narodu,
Z matki obcej, krew jego — dawne bohaterzy
A imię jego będzie czterdzieści i cztery.



Rys. 1. Udział rymów częstochowskich w Milionie (1 000 000) 44-rymowych podróbek z tekstu poematu Adama Mickiewicza „Pan Tadeusz” oraz przedział krytyczny na poziomie istotności 44 procent

Szczegóły opracowanego testu statystycznego opisane są w załączniku B. Na rysunku 1 przedstawiono natomiast rozkład Cz-punktów w 44-rymowych podpróbkach oraz granice przedziału krytycznego na poziomie istotności 44 procent. Utwory, dla których średni udział rymów częstochowskich leży poza tym przedziałem, czyli na krańcach albo „w ogonach” rozkładu, uznane są za statystycznie istotnie różne od wzorca, za jaki przyjęty został poemat „Pan Tadeusz”.

3. Częstochowszczyzna w poezji polskiej

Wprowadzenie Cz-punktów i zbadanie ich właściwości statystycznych pozwala na obliczenie stopnia częstochowszczyzny właściwego dla stylu różnych twórców. W tym celu utworzono zbiór (korpus) dokumentów zawierających dzieła polskich poetów kilku epok literackich. Oprócz utworów wieszczów Mickiewicza i Słowackiego uwzględniono twórczość noblistów, to jest Miłosza i Szymborskiej, poetów Kochanowskiego i Tuwima, autorów piosenek Osiecką i Kaczmarskiego oraz popularny zespół muzyczny Weekend. Bardziej szczegółowy opis korpusu „Częstochowa” znajduje się w tabeli 2.

Średni udział częstochowszczyzny obliczony dla stylu każdego z poetów przedstawiono na rysunku 2. Dokładność wyznaczenia tej średniej wzrasta wraz z liczbą przeanalizowanych rymów. Zobrazowano ją za pomocą poziomej kreski przedstawiającej przedział, który na 99 procent zawiera tę średnią. Dla każdego autora zaznaczono również górne

i dolne ograniczenie odpowiadające najbardziej częstochowskiemu i nieczęstochowskiemu sposobowi wyboru znaczników morfologicznych. Innymi słowy, wszystkie rymy, którym przypisano 50 Cz-punktów, zliczane są w pierwszym przypadku jako rymy częstochowskie $Cz = 100$, a w drugim jako nieczęstochowskie $Cz = 0$. Aby umożliwić wizualne przeprowadzenie testu częstochowskiego, odpowiadający mu zbiór krytyczny oznaczono pionowymi, przerywanymi liniami.

Tabela 2. Opis korpusu poezji „Częstochowa” analizowanego w niniejszej pracy

Autor	Utwory	Epoka literacka	Uwagi
Jan Kochanowski	<i>Fraszki</i>	renesans	mistrz Jan z Czarnolasu
Józef Baka	<i>Uwagi śmierci niechybnej</i>	barok	poeta metafizyczny
Adam Mickiewicz	<i>Pan Tadeusz</i>	romantyzm	wieszcz narodowy
Juliusz Słowacki	<i>Beniowski</i>	romantyzm	„wielkim poetą był”
autor nieznany	<i>Morskie opowieści</i>		popularna szanta
Maria Konopnicka	Poezja dla dzieci	pozytywizm	poetka
Julian Tuwim	<i>Bal w operze</i>	XX-lecie międzywojenne	Sejm RP ogłosił rok 2013 rokiem Tuwima
Czesław Miłosz	wiersze	współczesna	nagroda Nobla 1980
Wisława Szymborska	wiersze	współczesna	nagroda Nobla 1996
Agnieszka Osiecka	piosenki	współczesna	poetka, pisarka
Jacek Kaczmarski	poezja śpiewana	współczesna	poeta, kompozytor, bard „Solidarności”
Weekend	dyskografia	współczesna	gwiazdy disco polo

Najwyższy udział częstochowszczyzny w strofach analizowanych utworów przypadł dyskografii zespołu disco polo Weekend, chociaż i w tym przypadku co najmniej 30 procent rymów nie jest częstochowskich (Liszewski 2002):

Bum, bum, bum,
na faceta tak jak rum,
tak bardzo działa ruch damskiego ciała.
Bum, bum, bum,
facetowi tak jak rum
chodź ci pokażę swoje tatuaże.

Wyniki testu częstochowskiego sugerują, że poezja ta, pomimo znacznej popularności, nie jest przesadnie wyrafinowana. Analiza najnowszej płyty „Ona tańczy dla mnie” pokazuje jednak, że teksty piosenek grupy Weekend stają się porównywalne z „Morskimi opowieściami”, co świadczy o jej rozwoju artystycznym.

Wśród poetów, którzy już trafili do kanonu poezji polskiej, liderem w stosowaniu rymów gramatycznych okazał się ksiądz Józef Baka:

Cny młodziu, migdaliku,
Czerstwy rydzu, ślepowidzu (...)
Śmierć jak kot wpadnie w lot!

Jego barokowe strofy przez długi czas stawiane były za przykład grafomanii i złego smaku. Dopiero w XX wieku opinia o jego twórczości się zmieniła i doceniono go jako prekursora surrealizmu i poezji lingwistycznej. Wydaje się to wyjaśniać statystycznie istotne przekroczenie referencyjnego udziału częstochowszczyzny

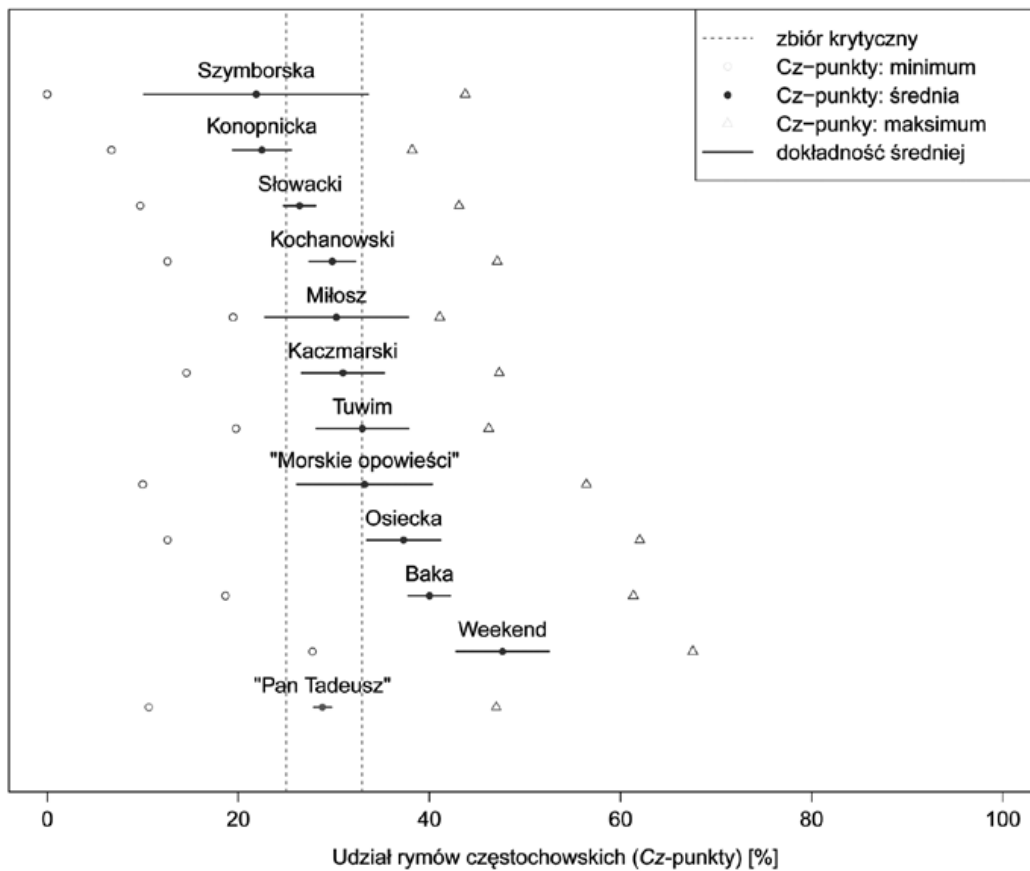
Kolejne miejsce w rankingu zajęła Agnieszka Osiecka, autorka wielu popularnych piosenek. Częstochowszczyzna w dziełach większości analizowanych poetów nie różni się istotnie od 44-rymowych podpróbek z poematu Adama Mickiewicza „Pan Tadeusz”. Co ciekawe, dzieła Kochanowskiego i Mickiewicza cechują się niemal identyczną liczbą Cz-punktów, i to pomimo dzielącego ich dwóch i pół wieku rozwoju języka. Być może zatem w poezji polskiej istnieje jakiś niezbędny udział rymów częstochowskich, wynikający z właściwości języka?

Słowacki układał swoje rymy bardzo starannie, uzyskując nieco mniejszą liczbę Cz-punktów. Fakt ten został dostrzeżony już przez Mickiewicza, który scharakteryzował poezję konkurenta jako:

gmach piękną architekturą stawiony, jak wzniosły kościół — ale w kościele Boga nie ma.

Powraca więc pytanie z „Ferdydurke” Gombrowicza: czy Słowacki wielkim poetą był? Architektura jego rymów rzeczywiście budzi podziw, jakkolwiek ocena przesłania utworów i ich głębi duchowej to odrębne zagadnienie, które wymaga przeprowadzenia analizy semantycznej. Ocena taka jest i na długo jeszcze pozostanie poza zasięgiem narzędzi analizy poezji wspomaganą komputerowo.

Najmniejszy udział rymów częstochowskich w swej twórczości poetyckiej uzyskała Maria Konopnicka w wierszach dla dzieci. Tak wyrafinowany warsztat literacki w twórczości przeznaczonej dla najmłodszych zasługuje na uznanie i podziw. Ocena laureatów nagrody Nobla, Miłosza i Szymborskiej przysparza trudności w związku z bardzo małą ilością rymów w ich poezji, a co za tym idzie — także z dużą niedokładnością wyznaczenia ich średniego udziału.



Rys. 2. Ranking wybranych poetów stosujących rymy gramatyczne (częstochowskie) w swojej twórczości

4. Podsumowanie

W tej pracy przedstawiono wyniki zautomatyzowanej oceny weny twórczej wybranych polskich poetów. Ograniczono ją jednak tylko do sprawności „warsztatowej” w zakresie jakości stosowanych rymów. Ich automatyczna identyfikacja oraz analiza morfologiczna pozwoliła na wyodrębnienie rymów gramatycznych zwanych częstochowskimi i określenie ich udziału w całości badanego dzieła. Poemat Adama Mickiewicza „Pan Tadeusz” posłużył jako tekst referencyjny, na podstawie którego opracowano statystyczny „test częstochowski” oraz utworzono ranking poetów, którego kryterium stanowi udział rymów gramatycznych w ich poezji. Uzyskane wyniki w zasadzie zgadzają się z istniejącą wiedzą w tym zakresie. Opracowanie zautomatyzowanego programu pozwalającego na wprowadzenie obiektywnych miar jakości warsztatu poetyckiego pozwala jednak dokładniej porównać styl rymowania poetów polskich. Tego typu analizy można wykonywać bez wsparcia komputerowego. Ich wiarygodność jest wówczas oparta również o semantyczną ocenę poszczególnych słów i zwrotów. Zadanie to jednak wymaga bardzo dużego nakładu pracy, a w przypadku analizy całej twórczości jakiegoś poety, np. trzeciego wieszca Zygmunta Krasińskiego, mogłoby się stać nużąco rozległe.

Informatyka wkracza w lingwistykę na wiele sposobów, poprzez poprawę błędów ortograficznych, interpunkcyjnych i gramatycznych, tezaury, automatyczne tłumaczenie tekstów na inne języki, czy też metody syntezy i rozpoznawania mowy. Proces ten będzie się stale rozwijał. Fakt, że coraz większa liczba utworów literackich jest dostępna w Internecie, tworzy możliwość przyspieszenia i obiektywizacji wielu analiz, zwłaszcza dotyczących kwestii formalnych. Przedstawiona praca mieści się w tym właśnie interdyscyplinarnym nurcie badawczym, gdyż jest próbą wykorzystania metod współczesnej informatyki do badań wybranych elementów struktury utworów poetyckich.

Załącznik A. Wykrywanie i analiza morfologiczna rymów

Literatura dotycząca automatycznego wykrywania rymów wciąż jest jeszcze uboga, być może dlatego, że CAP jest dopiero rodzącą się dziedziną badań. Genzel i in. (2010) pracowali nad przekładaniem wierszy na inne języki w taki sposób, aby zachować ich rytm i rymy. Szczegóły techniczne ich rozwiązania nie zostały jednak ujawnione, gdyż stanowią własność intelektualną prywatnej firmy. Hirjee i Brown (2010) scharakteryzowali ewolucję stylu rymowania w muzyce rap. Do automatycznego rozpoznawania rymów zastosowali zaawansowane narzędzia przygotowane na potrzeby bioinformatyki. Wyniki tej pracy pomagają w identyfikacji autorstwa, stylów muzycznych i tworzeniu systemów rekomendujących muzykę. Greene i in. (2010) posłużyli się analizą akcentów, rymów i wzorców rozmów do generowania poezji miłosnej. Wszystkie powyższe badania koncentrują się jednak na języku angielskim lub parze języków w przypadku tłumaczeń. Wykrywanie rymów dla języka polskiego wymaga zatem adaptacji metod znanych z literatury lub opracowania własnych rozwiązań.

Poezja jest zazwyczaj przechowywana w postaci tekstu, podczas gdy rymy zależą od jego wymowy. Pierwszym krokiem w wyszukiwaniu rymów jest zatem opracowanie reprezentacji dźwiękowej tekstu, którą w tej pracy oparto na fonemach zaczerpniętych z bazy CORPORA (Grocholewski 1997). W języku polskim zapis dość jednoznacznie określa wymowę, dlatego do jego konwersji na fonemy posłużono się prostym systemem regułowym. Ponadto, dla uproszczenia analizy, badaniu poddano jedynie rymy pomiędzy ostatnimi słowami każdej linii.

Ostatnie słowo każdego wersu porównywane było z ostatnimi słowami czterech poprzedzających wersów. Rymy wykrywano na podstawie podobieństwa pomiędzy trzema ostatnimi fonemami obu słów. W przybliżeniu odpowiada to sylabie, która według Śledzińskiego (2008) jest dobrą jednostką do analizy i syntezy mowy. Podobieństwo pomiędzy odpowiadającymi sobie parami fonemów zaczerpnięto z macierzy UPGMA. Wyznaczył ją Jakub Gałka w swojej pracy doktorskiej (2008), za pomocą analizy widmowej wykorzystującej 6-poziomową parametryzację dyskretnej tranformaty falkowej i falkę „dmey” dla pięciu różnych mówców. Podejście takie pozwala wykrywać zarówno rymy dokładne, jak i niedokładne. Zaproponowaną metodę można by usprawnić, biorąc pod uwagę koartykulację następujących po sobie fonemów (Grocholewski 1997; Śledziński 2008).

Do zbadania maestrii „warsztatowej” poetów tworzących w językach fleksyjnych konieczna jest analiza morfologiczna wykrytych rymów. W tym celu posłużono się anali-

zatozem *Morfeusz* (Saloni i in. 2011). *Morfeusz* analizuje słowa w oderwaniu od ich kontekstu, przez co listy możliwych znaczników, np. te przedstawione w tabeli 1, są często dosyć długie. Ich ujednoznaczenie wymagałoby przeprowadzenia analizy składniowej, która jest trudna w przypadku poezji i dla uproszczenia została pominięta w omawianym badaniu. Jej zastosowanie prawdopodobnie pozwoliłoby zwiększyć dokładność wyznaczania Cz-punktów, które w przedstawionym rozwiązaniu „ściągane” są do wartości 50 ze względu na kompromisowe rozwiązanie kwestii niejednoznaczności.

Załącznik B. Wyprowadzenie i właściwości formalne Cz-testu

Poemat referencyjny Adama Mickiewicza „Pan Tadeusz” zawiera ponad 4800 rymów (9850/2 wersów to 4925 par, ale w tekście występuje także wiele rymów potrójnych). Moc testów statystycznych wzrasta wraz z licznością próbki, co wyjaśnione jest np. w podręczniku Koronackiego i Mielniczuka (2009). Z tej przyczyny zdecydowano się unormować długość analizowanego wiersza do 44 kolejnych rymów. Aby zbadać rozrzut Cz-punktów wynikający z takiego ograniczenia długości, wylosowano ze zwracaniem 1 000 000 takich podpróbek, przy czym 1 000 000, czyli Milijon, to kolejna stała zaproponowana w „Wielkiej improwizacji” (Mickiewicz 1832):

Nazywam się Milijon — bo za miliony kocham i cierpię katusze.

W ten sposób uzyskano rozkład Cz-punktów dla 44-rymowych prób z „Pana Tadeusza”, którego prawdopodobieństwo zobrazowane jest na rysunku 1. Na podstawie tego rozkładu można utworzyć nieparametryczny test statystyczny na poziomie istotności α poprzez odcięcie $\alpha/2$ procent prawdopodobieństwa z każdego z ogonów (krańców) rozkładu statystycznego. Jeśli średni udział rymów gramatycznych dla nowej obserwacji, np. nowego tomiku poezji, wypadnie w któryś z odciętych obszarów, to można stwierdzić, że zaobserwowane różnice są statystycznie istotne. Wybór poziomu istotności jest zawsze kwestią arbitralną, zaś w tej pracy przyjęto wartość $\alpha = 44$ procent.

Okazuje się, że „test częstochoowski” można bardzo dobrze przybliżyć testem *t*-Studenta. Zgodnie z centralnym twierdzeniem granicznym wraz ze wzrostem liczności próbki rozkład średniej niezależnie wylosowanych rymów zbiega do rozkładu normalnego. Aby stwierdzić, czy 44 jest dostatecznie dużą liczbą, sprawdzono 1 000 00 niezależnych prób z tekstu „Pana Tadeusza” za pomocą siedmiu testów normalności. Żaden z nich nie zdołał odrzucić hipotezy zerowej na poziomie ufności ponad 99,9 procent. Oznacza to, że można modelować rozkład średniej z 44-rymowej próbki za pomocą rozkładu normalnego, którego wartość oczekiwana μ jest równa średniej wartości Cz-punktów dla całego dzieła, a odchylenie standardowe σ można obliczyć z następującej zależności:

$$\sigma = \sqrt{\frac{p_0 \cdot 0^2 + p_{50} \cdot 50^2 + p_{100} \cdot 100^2 - \mu^2}{n}} = \sqrt{\frac{2500p_{50} + 10000p_{100} - \mu^2}{n}} \quad (1)$$

Licznik równania (1) opisuje odchylenie stan dardowe rozkładu Cz-punktów w całym utworze. Rozkład ten jest trzypunktowy, dlatego można go obliczyć z definicji, poprzez wyznaczenie udziałów p_0 , p_{50} i p_{100} rymów przypadających na każdą kategorię punktów „częstochowszczyzny”.

$$p_i = \frac{\text{liczba rymów, dla których } Cz = i}{\text{ogólna liczba rymów}}, \quad \text{dla } i = 0, 50, 100. \quad (2)$$

Pierwiastek z 44 w mianowniku równania (1) odzwierciedla fakt, że obliczana jest średnia z próbki czterdziestu czterech niezależnie wylosowanych rymów.

Bibliografia

- Anonimowy recenzent, 2010, Paper review in verse. Recenzja dostępna online http://research.google.com/archive/papers/review_in_verse.html [17.07.2013 r.].
- Baka J., 1766, Uwagi o śmierci niechybnej, wszystkim pospolitej, wierszem wyrażone, a sumptem Jmci pana Ksawerego Stephaniego, obywatela miasta J.K.M. Wilna, do druku na użytek duchowny podane w roku 1766.
- Galeria, 2012, Częstochowski magazyn literacki Literackiego Towarzystwa Wzajemnej Adoracji Li-TWA, nr 24. Magazyn dostępny online <http://www.litwa.czyst.pl/archiwum-galerii> [17.07.2013 r.].
- Gałka J., 2008, Optymalizacja parametryzacji sygnału w aspekcie rozpoznawania mowy polskiej, rozprawa doktorska, Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie.
- Genzel D., Uszkoreit J., Och, F., 2010, “Poetic” statistical machine translation: Rhyme and meter, w materiałach konferencji 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP’10, s. 158–166, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Gombrowicz W., 1937, *Ferdydurke*.
- Greene E., Bodrumlu T., Knight K., 2010, Automatic analysis of rhythmic poetry with applications to generation and translation, w materiałach konferencji 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP’10, s. 524–533, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Grochowski S., 1997, Corpora — speech database for Polish diphones, w materiałach konferencji EUROSPEECH.
- Hirjee H., Brown D.G., 2010, Using automated rhyme detection to characterize rhyming style in rap music, *Empirical Musicology Review*, nr 5(4), s. 121–145.
- Koronacki J., Mielniczuk J., 2009, *Statystyka dla studentów kierunków technicznych i przyrodniczych*, Warszawa.
- Liszewski R., 2002, *A Ja To Lubię (Bum Bum)*, piosenka grupy Weekend z płyty „Playboy” wydanej przez Green Star.
- Mickiewicz A., 1832, *Wielka Improwizacja*, [w:] *Dziady część III*.

- 1834, *Pan Tadeusz*, czyli Ostatni zajazd na Litwie: historia szlachecka z roku 1811 i 1812 we dwunastu księgach wierszem.
- Piersiak T., 2008, „pana Częstochowy” nie możemy być dumni, *Gazeta Wyborcza*, 23 października 2008 r.
- Saloni Z., Gruszczyński W., Woliński M. i Wołosz R., 2011, *Analizator morfologiczny Morfeusz*.
- Śledziński D., 2008, Fonetyczno-akustyczna analiza struktury sylaby w języku polskim na potrzeby technologii mowy, *Investigationes Linguisticae*, nr 16, s. 219–240.
- Woliński, M., 2003, 2003, System znaczników morfosyntaktycznych w korpusie IPI PAN, *Polonica XXII–XXIII*, s. 39–55.

SUMMARY

Grammatical Rhymes in Polish Poetry: a Quantitative Analysis

Analysis and interpretation of poetry is based on qualitative features of its text such as its semantics or means of expression as well as on general knowledge about the author and artistic period. Recent advances in automatic text processing allow for performing quantitative analysis of large sets (corpora) of poetry, whose results may facilitate assessment of linguistic capabilities of its author or in other words, his poetic mastery. This contribution presents a method of calculating the share of grammatical rhymes in Polish poetry, known as “Częstochowa” rhymes. This value is used to create a ranking of both historic and contemporary Polish poets based on technical quality of their rhymes. Comparative study and statistical analysis is developed using *Pan Tadeusz* by Adam Mickiewicz as a reference poem for Polish poetry. Assessment of technical mastery is the first step towards introduction of objective measures of poetic vein, which is one of the urging challenges of Computer-Aided Poetry.