

Annales Universitatis Paedagogicae Cracoviensis

Studia Poetica 12 (2024)

ISSN 2353-4583

e-ISSN 2449-7401

DOI 10.24917/23534583.12.7

Aleksandra Rykowska

Uniwersytet Jagielloński w Krakowie

ORCID: 0009-0003-6337-793X

Jak i po co automatyzować analizę wersologiczną w polskich badaniach literackich?

Wprowadzenie do tematyki

Dzięki badaniom, którym przyświecał strukturalny duch, rozwijanym m.in. przez Marię Dłuską czy Lucyllę Pszczołowską, analiza wersologiczna w polskim literaturoznawstwie była jedną z prężniej rozwijanych dziedzin poetyki. Ostatnimi czasy badania skupione wokół wersyfikacji są często marginalizowane, a czasem i zupełnie pomijane w nowszych odczytaniach poezji, zarówno dawnej, jak i współczesnej. Warto przypomnieć więc słowa przytoczone we wstępie najważniejszego zbioru interpretacji polskiej poezji wydanego pod redakcją Janusza Sławińskiego oraz Jana Prokopa, że „sztuka interpretacji przeciwstawia się mniemaniom, które za ideał odbioru poezji uznają odbiór bezrefleksyjny, pozbawiony nastawień analitycznych. Wychodzi z założenia, że dopiero utwór właściwie zrozumiany, uchwycony w bogactwie swoich immanentnych i zewnętrznych powiązań, jest obiektem naprawdę przeżytym przez odbiorcę”¹.

Nie sposób zapomnieć o dokładnej analizie, by zaprezentować dogłębną, dobrze wykonaną interpretację utworu literackiego. I, mimo że strukturalistyczne analizy wykonywane w drugiej połowie ubiegłego stulecia mogą odstraszać odbiorców swoim formalizmem i skłonnością do wyraźnej przewagi aspektu analitycznego nad refleksyjnym i interpretacyjnym, obecnie polskie badania literackie powinny postarać się poszukać złotego środka, gdzie oba te elementy czytania poezji nie

¹ *Liryka polska. Interpretacje*, red. J. Prokop, J. Sławiński, słowo/obraz terytoria, Gdańsk 2001, s. 6.

będą się ze sobą kłócić, a wzajemnie uzupełniać. Wnioski analityczne staną się przyczynkiem i potwierdzeniem pewnych tez interpretacyjnych, a wykładniki formalne analizy będą współgrać z osobistymi wnioskami i przeżyciami interpretatora.

Nie należy ukrywać, że humanistów, szczególnie tych zajmujących się poezją, formalna, matematyczna i statystyczna analiza mogą odstraszać. Zapominamy przy tym, że metryka, rytm, wersyfikacja, to nic innego jak... liczenie. Do jakich wniosków można więc dojść, gdy policzymy różne elementy utworu poetyckiego, zaczynając od strof, wersów, poszczególnych wyrazów czy metafor, a kończąc na językowych elementach składających się na tekst wiersza – częściach mowy, a nawet pojedynczych fonemach? Być może odkryjemy wtedy lingwistyczne dowody na potwierdzenie subiektywnych odczuć odbiorcy wiersza. Taka perspektywa wydaje się bardzo obiecująca, jednak równocześnie monotonna i pomyłkogeniczna. Nie dziwi więc, że przeprowadzona w ten sposób mistrzowska analiza wiersza Kazimierza Wierzyńskiego o incipicie *Gdzie nie posieją mnie...* w wykonaniu Marii Dłuskiej², pochodząca ze wspomnianego tomu interpretacji polskiej liryki, jest jedynym tego rodzaju badaniem przeprowadzonym w historii polskich badań literackich.

W takich momentach polskim literaturoznawcom i historykom literatury z pomocą mogą przyjść nowe technologie. Jeżeli potrafimy coś sklasyfikować i policzyć, a do tego określić szczegółowe kryteria i reguły, według których miałyby to zrobić człowiek – dlaczego nie miałyby tego wykonać za nas komputer? W ten sposób eliminujemy ryzyko popełnienia błędu spowodowanego mniejszym skupieniem uwagi czy zmęczeniem, dodatkowo mając pewność, że wyniki analizy są przeprowadzone w całości poprawnie. Z chęci przywrócenia sposobu badań zaproponowanego przez Dłuską, który to może być ważnym elementem przyszłych badań nad ikoniznością języka i stylometrią krótszych utworów poetyckich, powstał program SylLab. Niniejszy artykuł przedstawia sposób działania algorytmu omawianego programu oraz propozycję wykorzystania wyników uzyskanych w czasie przeprowadzonej w nim analizy wersologicznej i fonetycznej wiersza.

Cel pracy

Celem niniejszego artykułu jest przeprowadzenie dowodu, na konkretnym przykładzie, że technologie komputerowe, odpowiednio przemyślane i bazujące na dobrze znanych z przeszłości rozwiązaniach, mogą niewątpliwie wzbogacić interpretacje historycznoliterackie. O ile sam komputer (jak i metoda badawcza jako taka) nie zastąpi intuicji oraz nie wyciągnie za badacza wniosków, to może być on bardzo użytecznym narzędziem do usprawnienia analizy. Komputer można stosunkowo szybko dostosować do rozwiązania konkretnego problemu, pisząc stosowne oprogramowanie. Wykona on wtedy analizę szybciej i dokładniej niż człowiek,

² M. Dłuska, *Kazimierz Wierzyński «Gdzie nie posieją mnie»*, w: *Liryka polska. Interpretacje*, red. J. Prokop, J. Sławiński, słowo/obraz terytoria, Gdańsk 2001, s. 293–319.

jednocześnie wizualizując i porównując wyniki badań. Wydaje się, że efektywne połączenie umiejętności programistycznych i językowych może skutkować nową jakością w dziedzinie badań humanistycznych.

W artykule przedstawiono więc podstawowe informacje, które pozwalają połączyć analizę wersologiczną utworu z badaniami nad ikonicznością dźwięków mowy, co dalej powiązane być może z automatycznym rozpoznawaniem nastroju danego utworu poetyckiego. Zaprezentowane zostało działanie autorskiego narzędzia opracowanego na podstawie artykułu Marii Dłuskiej – artykuł zawiera więc opis algorytmu, który może sprawić, że zapomniana metoda, użyta jedynie raz przez Dłuską, zostanie na nowo zauważona przez badaczy literatury.

Algorytm działania Marii Dłuskiej

W swojej interpretacji wiersza Wierzyńskiego Maria Dłuska postuluje, że na podstawie porównania badanego utworu z minikorpusem referencyjnym wierszy powstałych w podobnym okresie literackim, posiadających podobną kompozycję oraz strukturę rytmiczną, można określić nastrój wiersza (pozytywny/negatywny), obliczając stosunki głosek wysokich i niskich w utworze i zestawiając ze sobą wyniki procentowe obliczeń dla różnych wierszy. Teza ta wydaje się niezwykle ciekawa głównie z tego względu, że zakłada połączenie między warstwą dźwiękową utworu, a jego nastrojem, co dalej można rozumieć jako postulowanie ukrytej ikoniczności czy symbolizmu dźwięków mowy. Niestety w omawianym tekście Dłuskiej próżno szukać lingwistycznych dowodów na słuszność postulowanych tez, stąd w niniejszym artykule nie zabraknie przytoczenia badań, które mogłyby przemawiać za słusznością nowatorskich na tamte czasy przeczuc badaczki. Przed wprowadzeniem do działania samego programu należy więc wytłumaczyć, dlaczego powinno się zestawiać ze sobą wiersze o tym samym schemacie rytmicznym oraz jakie właściwości wysokich dźwięków mowy mogą przemawiać za ich pozytywnym odbiorem słuchowym.

Co kryje rytm?

Rytm wiersza jest dlań kluczowy – jest podstawowym czynnikiem organizującym utwór poetycki. Już przy pierwszym głośnym czytaniu wiersza posiadającego mocno zarysowaną i regularną strukturę rytmiczną można zauważyć, że układ akcentów odpowiada w nim za to, że tekst brzmi skocznie, radośnie czy monotonicznie i ponuro. Niektórzy również będą doszukiwać się śladów pewnego odzwierciedlenia rytmu wykonywania niektórych czynności właśnie w tym, jak autor postanowił wykorzystać pozycję akcentowanych sylab, co może nas skłaniać ku tezie o ikoniczności rytmu wiersza. Teza ta jest szczególnie popularna w rosyjskojęzycznych kręgach poetyckich, gdzie największym zainteresowaniem cieszy się pen-

tametr trochaiczny, przypominający rytm ludzkiego chodu³. Zauważono również, że ten układ wersologiczny pojawia się skrupulatnie w wierszach, w których powtarza się jako główny wątek motyw drogi i podróży. To wszystko zapoczątkowało z kolei tezę o nazwie »semantic halo of verse«, według której rytm wiersza jest tym, co narzuca w pewien sposób temat wiersza, być może właśnie ze względu na częściową ikoniczność rytmu i fakt, że przypomina odbiorcy o pewnych aspektach rzeczywistości, które odczuwamy właśnie przez miarowe powtarzanie pewnych ruchów, dźwięków itd⁴.

Metody komputerowe nie tylko mogą usprawnić analizę wersologiczną, ale również potwierdzać pewne teorie, takie jak ta przedstawiona w tym artykule. W ciągu ostatnich lat przeprowadzono eksperymenty, które potwierdziły efekt »semantic halo« dla wierszy nie tylko rosyjskojęzycznych, ale także dla innych języków indoeuropejskich, m.in. dla języka czeskiego, co jest o tyle ważne w kontekście polskich badań literackich, że oba te języki, pochodzące z tej samej rodziny językowej i będące blisko spokrewnione, wykazują podobne cechy⁵. Można więc zaryzykować stwierdzenie, że opisywany efekt prawdopodobnie występuje także w polskiej poezji.

Jeżeli więc rytm narzuca wierszom pewną semantykę i temat, to co może te wiersze różnicować na niższym poziomie reprezentacji językowej? Odpowiedź to: głoski i ich cechy fonetyczne.

Symbolizm dźwięków mowy

Dyskusja na temat ikoniczności języka i symbolizmu poszczególnych dźwięków mowy cofa się jeszcze do czasów Platona i pierwszej udokumentowanej refleksji językoznawczej na kontynencie europejskim, zapisanej na kartach *Kratylosa*. Już w antyku naturaliści przegrali dyskusję, jakoby mowa miała być odzwierciedleniem otaczającego świata, a jej koniec przypieczętował wydany w 1916 roku *Kurs językoznawstwa ogólnego* opracowany na podstawie wykładów Ferdnina de Saussure'a, w którym znajduje się obowiązująca do dzisiaj definicja języka jako zbioru arbitralnych znaków⁶.

Mimo pewnego tabu, które w językoznawstwie objęło problem ikoniczności, temat ten powracał w przeciągu ostatnich lat i z największą siłą wrócił w momencie tzw. zwrotu kognitywnego, od kiedy to obserwuje się zdecydowane zwiększe-

³ M. Trunin, *Towards the concept of semantic halo*, „Studia Metrica et Poetica” 2017, t. 4, nr 2, s. 41–66.

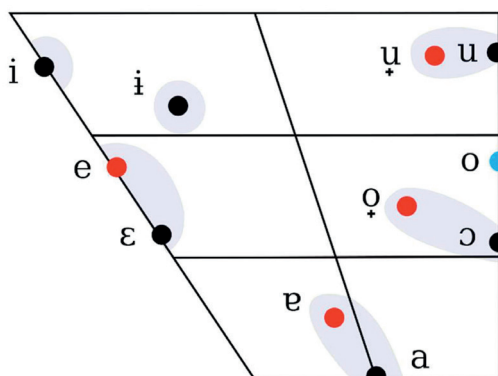
⁴ Tamże.

⁵ A. Šeja, P. Plecháč, A. Lassche, *Semantics of European poetry is shaped by conservative forces: The relationship between poetic meter and meaning in accentual-syllabic verse*, „PLoS ONE” 2022, nr 17 s. 1–17.

⁶ F. de Saussure, *Kurs językoznawstwa ogólnego*, przeł. K. Kasprzyk, Warszawa 2007, s. 40–41.

nie badań dotyczących ikoniczności. Najpopularniejszym badaniem, które należy przywołać, jest eksperyment dotyczący tzw. efektu buba-kiki⁷. Wtedy zauważono, że dźwięki [i] oraz [a] są wyraźnie nacechowane, a wspólne odczucia dotyczące tych dźwięków są powtarzalne wśród rodzimych użytkowników różnych, nawet niespokrewnionych, języków. W ciągu kolejnych lat udowodniano, że charakterystyka tych dwóch fonemów odpowiada również za to, że postrzegamy niektóre dźwięki jako „ciemniejsze”, inne za „jaśniejsze”, lub też mniej lub bardziej pozytywne i negatywne⁸.

Teorią, która podnoszona jest najczęściej (i którą wzięła pod uwagę również Maria Dłuska) jest ta, w której za bardziej lub mniej pozytywny odbiór poszczególnych dźwięków mowy odpowiada pozycja języka w jamie ustnej w czasie artykulacji. Im wyżej znajduje się język, tym dźwięk jest postrzegany jako bardziej pozytywny. Dla języka polskiego pozycję tego artykulatora przedstawić można na czworokącie samogłoskowym (Rys. 1).



Rys. 1. Czworokąt samogłoskowy dla języka polskiego

Źródło: M. Wiśniewski, *Zarys fonetyki i fonologii współczesnego języka polskiego*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń 2007.

Za wysokie są więc uważane samogłoski [i], [i] oraz [u]. [e] jest samogłoską średnią, [ɔ] oraz [a] są niskie. Ponieważ głoska [i] w języku polskim zmiękcza poprzedzające spółgłoski, one również są traktowane w opisywanej w niniejszym tekście metodzie, jak dźwięki wysokie (dla spółgłosek należy wskazać opozycję

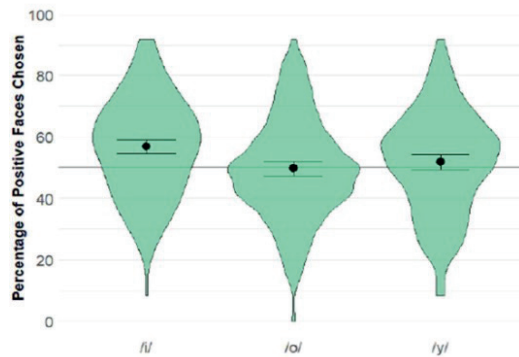
⁷ V. Ramachandran: *Neuronauka o podstawach człowieczeństwa*, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 2012, s. 129.

⁸ J.O. Svantesson, *Sound symbolism: the role of word sound in meaning*, „WIREs Cogn Sci” 2017, nr 8, s. 1–12; J.B. Nuckolls, *The case for sound symbolism*, „Annu. Rev. Anthropol.” 1999, nr 28, s. 225–252.

miękka-twarda, gdzie głoska miękka charakteryzuje się podniesieniem języka w czasie artykulacji).

Takie traktowanie dźwięków mowy jest zgodne m.in. z teorią muzyki, w której podobne nacechowanie przypisywane jest dźwiękom wyższym i niższym, czy też w tonacji durowej i molowej⁹. Eksperymentalnie takie podejście zostało potwierdzone w badaniu przeprowadzonym na niemieckojęzycznych mówcach, którzy wysłuchali logatomów zawierających poszczególne dźwięki samogłoskowe i mieli oceniać ich nacechowanie. Dźwięk [i] był oceniany jako bardziej pozytywny od innych dźwięków (Rys. 2).

Results from Experiment 4



Rys. 2. Wyniki eksperymentu dot. walencji samogłosek

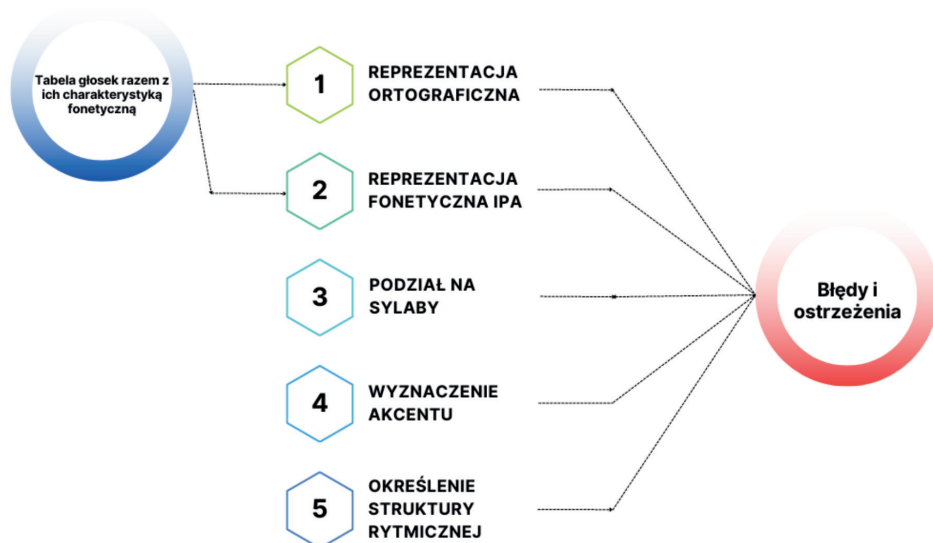
Źródło: A. Körner, R. Rummer, *Articulation contributes to valence sound symbolism*, „Journal of Experimental Psychology General” 2021, nr 151(5), s. 16.

Mimo że przedstawione powyżej badania nie są argumentem jednoznacznie potwierdzającym tezę o ikonizacji wymienionych dźwięków mowy, są one jednak potwierdzeniem słuszności przypuszczeń Marii Dłuskiej oraz zachętą do przeprowadzania kolejnych badań dotyczących tego problemu. Analiza polskiej poezji wykonana z użyciem programu Syllab może więc przyczynić się do dodania kolejnej cegiełki do badań ikonizacji dźwięków mowy.

⁹ I. Iwańska, *Jasność i ciemność, czyli tonacje krzyżykowe i bemolowe*, „Teoria muzyki” 2018, nr 12, s. 105–125.

Reguły automatyzacji analizy wersologicznej

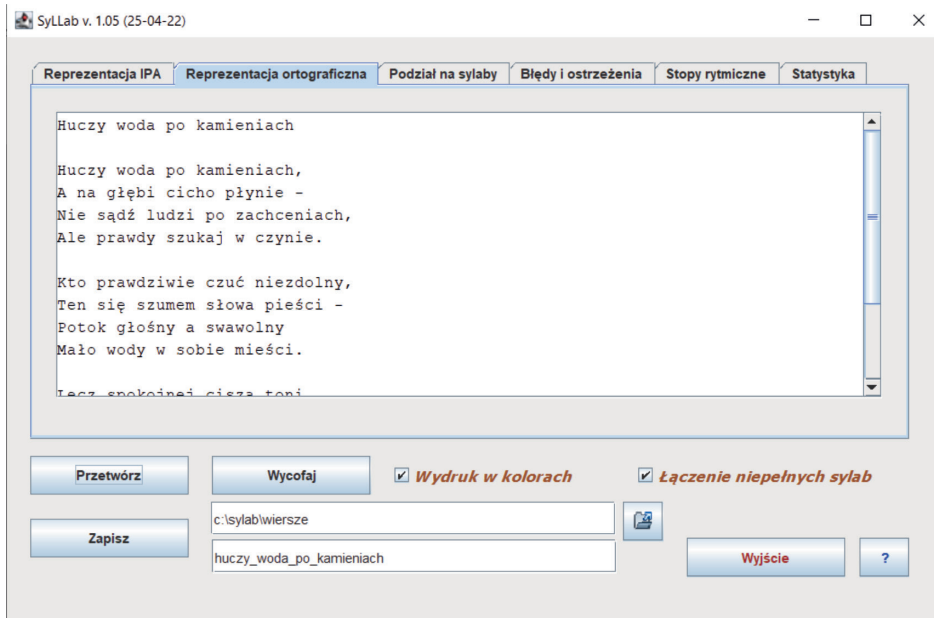
Dokładne prześledzenie analizy Marii Dłuskiej pomaga w ustaleniu algorytmu działania, który następnie można zamienić w kod programu opierającego się na regułach językowych. W celu znalezienia wiersza o tej samej strukturze rytmicznej co utwór, który ma podlegać analizie, należy wpierw ustalić strukturę rytmiczną tego utworu oraz innych, które wejdą w skład korpusu referencyjnego. Aby mieć dostęp do tych danych, program musi wykonać pięć niezbędnych kroków, które są przedstawione na poniższym schemacie (Rys. 3).



Rys. 3. Schemat działania programu Syllab

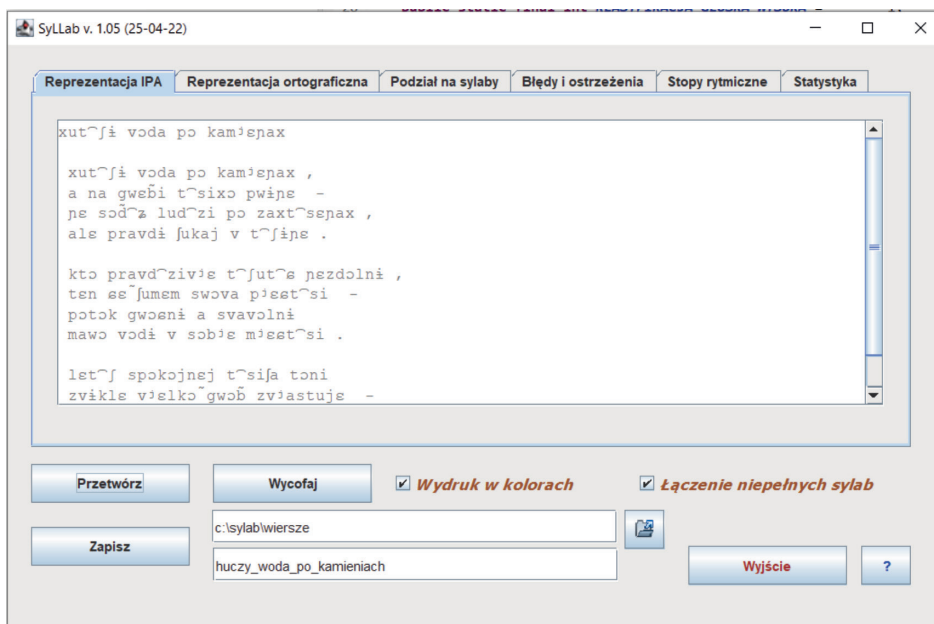
Źródło: opracowanie własne.

Narzędzie jako tekst wejściowy obsługuje zarówno tekst w reprezentacji ortograficznej, która jest następnie zamieniana na reprezentację fonetyczną (zapisaną w międzynarodowym alfabecie fonetycznym IPA) lub może analizować już przetaskrybowany tekst. Program wykonuje transkrypcję opierając się na zestawie reguł, które zostały wprowadzone z myślą o przeznaczeniu wersyfikacyjnym programu (Rys. 4-5). Z tego względu wykonana transkrypcja nie jest idealną reprezentacją fonetyczną – nie uwzględnia np. asymilacji dźwięczności czy upodobnień międzywyrazowych, ponieważ cecha dźwięczności nie wpływa w żaden sposób na dalsze etapy analizy.



Rys. 4. Tekst analizowanego wiersza

Źródło: opracowanie własne.



Rys. 5. Automatycznie wygenerowana transkrypcja fonetyczna IPA

Źródło: opracowanie własne.

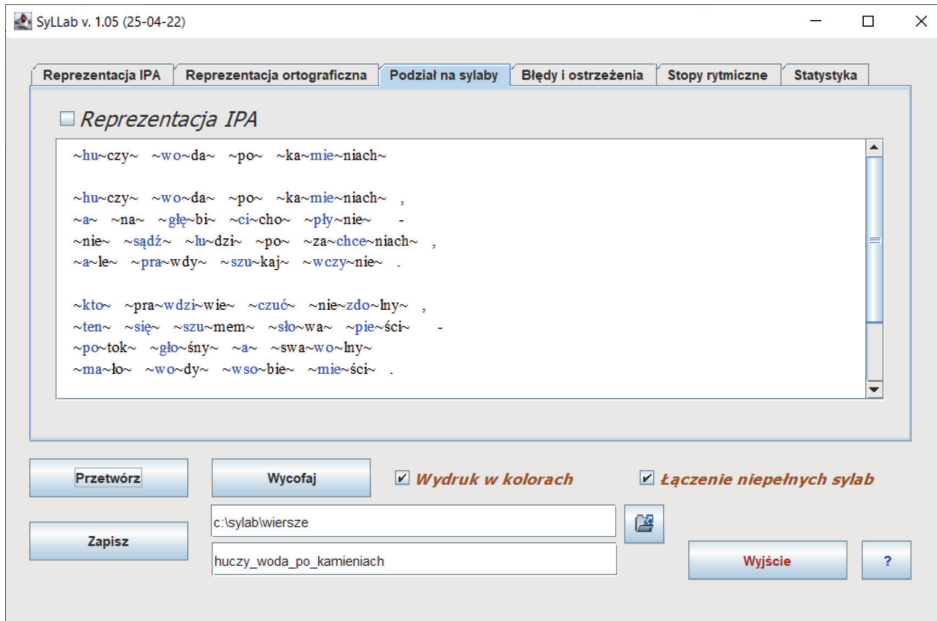
Uproszczona reprezentacja fonetyczna jest niezbędna do poprawnego wykonania następnego kroku, jakim jest podział na sylaby słów wchodzących w skład każdego wersu utworu. Podział ten jest wykonany na podstawie teorii o hierarchii sonorności, która jest bardzo silną tendencją w języku polskim. Granica sylaby przebiega w miejscu, w którym wartość sonorności głosek gwałtownie spada między dwoma różnymi dźwiękami. Wartości sonorności dla poszczególnych głosek zostały przedstawione w Tabeli 1.

Tabela 1. Klasy sonorności głosek

| Klasa głosek | Głoski wchodzące do klasy | Sonorność |
|---|---|-----------|
| Samogłoski | [a, ε, ɔ, u, i, ɨ, ɛ̃, ɔ̃] | 6 |
| Półsamogłoski | [j, w] | 5 |
| Spółgłoski płynne | [l, l̥, r, r̥] | 4 |
| Spółgłoski nosowe | [m, m̥, n, n̥, ŋ, ŋ̥] | 3 |
| Spółgłoski szczelinowe | [f, f̥, v, v̥, s, s̥, z, z̥, ʃ, ʃ̥, x, x̥] | 2 |
| Spółgłoski zwarte: zwarto-wybuchowe, zwarto-szczelinowe | [t̪s, t̪s̥, t̪e, t̪e̥, t̪ʃ, t̪ʃ̥, d̪z, d̪z̥, d̪ʒ, d̪ʒ̥, p, p̥, b, b̥, t, t̥, d, d̥, k, c, g, ʝ] | 1 |

Oprac. na podstawie: D. Śledziński, *Tworzenie reguł dla programu dzielącego tekst w języku polskim na sylaby*, „Biuletyn Polskiego Towarzystwa Językoznawczego” 2016, t. 72, s. 151–161.

Granica między sylabami jest oznaczona symbolem tyldy. Na tym etapie program ustala również, które sylaby są akcentowane według zasady, że akcent w języku polskim jest stały pod względem fonetycznym i pada na przedostatnią sylabę danego wyrazu (akcent paroksytoniczny). Ustalona została również lista wyjątków, kiedy akcent wyrazowy jest proparoksytoniczny, oksytoniczny, jak i lista wyrazów, które nie posiadają własnego akcentu (klityki). Program traktuje je jak jeden wyraz, razem ze słowem, które następuje po nich (Rys. 6).



Rys. 6. Podział tekstu na sylaby

Źródło: opracowanie własne.

Z takiej reprezentacji program pozwala wygenerować również schemat rytmiczny utworu, który można zapisać w formacie *.txt* oraz *.htm* do dalszych analiz (Rys. 7).

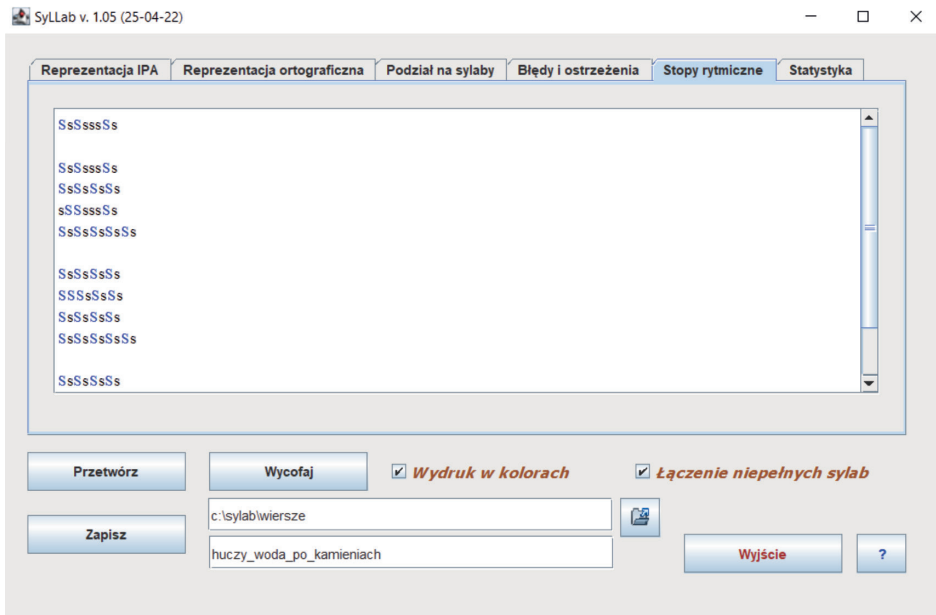
Narzędzie działające w ten sposób może okazać się pomocne przy tworzeniu korpusu polskiej poezji, który do tej pory nie został jeszcze niestety wykonany przez żadną instytucję naukową. Warto jednak zainspirować się m.in. czeskim korpusem poezji, który jest anotowany chociażby wersologicznie¹⁰.

Funkcjonalność programu SylLab nie kończy się jednak w tym miejscu. Ponieważ program operuje na liście z cechami głosek, za jego pomocą przeprowadzić można także dalsze etapy analizy inspirowanej przez badanie Marii Dłuskiej. Ostatnia zakładka widoczna w graficznym interfejsie użytkownika umożliwia podgląd statystyki różnych typów głosek, które znajdują się w utworze (Rys. 8-9).

Wyniki te można eksportować do arkusza *.csv*, który następnie może służyć do dalszych analiz statystycznych. Umożliwia on również prześledzenie, wers po wersie, wystąpień wszystkich typów głosek, gdyby badacz nie potrzebował analizować całego utworu, a jedynie jego fragment (Rys. 10).

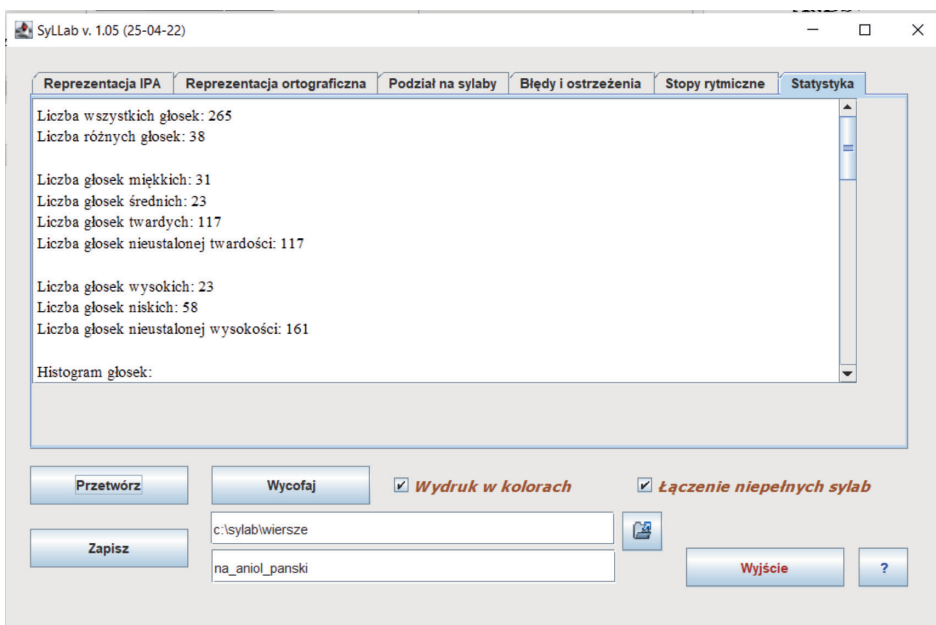
¹⁰ P. Plecháč, R. Kolár, *The Corpus of Czech Verse*, „*Studia Metrica et Poetica*” 2015, nr 1, s. 107–118.

Jak i po co automatyzować analizę wersologiczną w polskich badaniach literackich? [123]



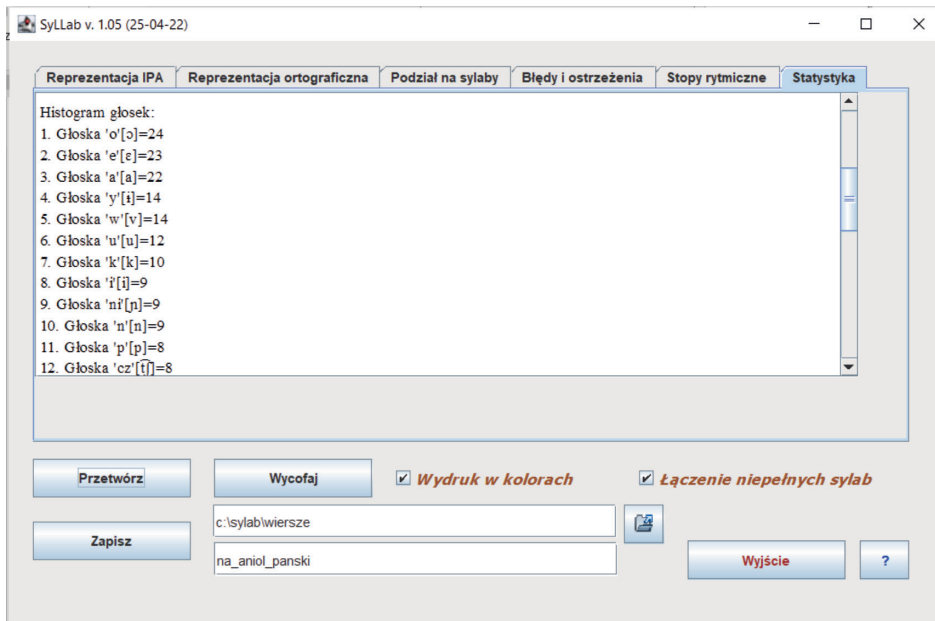
Rys. 7. Reprezentacja rytmu wiersza

Źródło: opracowanie własne.



Rys. 8. Liczbowe podsumowanie wystąpień kategorii głosek

Źródło: opracowanie własne.



Rys. 9. Histogram głosek

Źródło: opracowanie własne.

| 1 | Statystyka głosek: | Wers 1 | Wers 3 | Wers 4 | Wers 5 | Wers 6 | Wers 8 | Wers 9 | Wers 10 | Wers 11 | Wers 13 | Wers 14 | Wers 15 | Wers 16 | Suma głosek w wiersze |
|----|------------------------|--------|--------|--------|--------|--------|--------|--------|---------|---------|---------|---------|---------|---------|-----------------------|
| 2 | Głoska | | | | | | | | | | | | | | |
| 3 | u[ɔ]=24 | 2 | 2 | 1 | 1 | 2 | 1 | 4 | 3 | 3 | | | | | 23 |
| 4 | e[ɛ]=23 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 3 | 1 | 1 | 21 |
| 5 | a[a]=22 | 3 | 3 | 2 | 2 | 3 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | | 22 |
| 6 | y[ɨ]=14 | 1 | 1 | 1 | | 2 | 1 | | 2 | 1 | | | 1 | 1 | 12 |
| 7 | w[v]=14 | 1 | 1 | | 2 | 1 | 1 | 2 | 2 | | | | 1 | 2 | 14 |
| 8 | u[u]=12 | 1 | 1 | | 1 | 1 | 1 | 1 | | | | | 1 | 2 | 11 |
| 9 | k[k]=10 | 1 | 1 | | | 1 | 1 | | | 1 | | 1 | 2 | 2 | 10 |
| 10 | i[i]=9 | | | 1 | | 1 | | | | | | 2 | | | 4 |
| 11 | n[n]=9 | | | 1 | | | 1 | 1 | 2 | | | 2 | | 1 | 8 |
| 12 | m[m]=9 | 1 | 1 | 1 | 2 | 1 | 1 | | | | | | 2 | | 9 |
| 13 | s[s]=8 | | | | 1 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| 14 | t[t]=8 | | | | | | | 1 | 1 | 1 | | 1 | 1 | 2 | 8 |
| 15 | p[p]=8 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | 1 | | | 8 |
| 16 | cz[tʃ]=8 | 1 | 1 | | | 1 | 1 | | | | | 1 | 1 | 1 | 7 |
| 17 | ʃ[ʃ]=7 | | | 2 | | | | 1 | 1 | 1 | | | | 1 | 7 |
| 18 | ʎ[ɲ]=7 | | | | 1 | 1 | 1 | | | 1 | | 1 | 2 | | 7 |
| 19 | ʎ[ɲ]=7 | 1 | 1 | | | 1 | 1 | | | | 1 | | | 1 | 6 |
| 20 | h[h]=6 | | 1 | 1 | | | | | | | | | | | 2 |
| 21 | ʦ[ʦ]=5 | | | 1 | 1 | | | | | | | 1 | | | 3 |
| 22 | ʒ[ʒ]=5 | | | | | 1 | | 1 | | | 1 | | | 1 | 4 |
| 23 | ʒ[ʒ]=5 | | | | | | 1 | | | | 2 | | 2 | 1 | 5 |
| 24 | w[v]=4 | | | | | | | | | | | | 2 | 1 | 4 |
| 25 | ʃ[ʃ]=4 | | | | 1 | 1 | | | | | | 2 | | | 4 |
| 26 | r[r]=4 | | | | | 1 | 1 | | | | | | 2 | | 4 |
| 27 | ʂ[ʂ]=4 | | | 1 | | | | | 1 | | | | 1 | 2 | 4 |
| 28 | m[m]=3 | 1 | 1 | | | | | | | 1 | | | | | 3 |
| 29 | ʒ[ʒ]=3 | | | | 1 | | | | | | | 2 | | | 3 |
| 30 | ʎ[ɲ]=3 | | | | | | | 1 | | | | | | | 1 |
| 31 | m[m]=3 | | | | | | | 2 | | 1 | | | | | 3 |
| 32 | b[b]=3 | | | | | | | | 1 | | | 1 | | 1 | 2 |
| 33 | ʦ[ʦ]=3 | | | 1 | | | | | | | | | | 1 | 3 |
| 34 | ʦ[ʦ]=2 | | | | | | | 1 | | 1 | | | | | 2 |
| 35 | h[h]=2 | 1 | 1 | | | | | | | | | | | | 2 |
| 36 | ʂ[ʂ]=2 | | | | 1 | | | | | | | | | | 2 |
| 37 | d[ɗ]=2 | | | 1 | | | | | | 1 | | | | | 2 |
| 38 | ʂ[ʂ]=1 | | | | | | | 1 | | | | | | | 1 |
| 39 | ʂ[ʂ]=1 | | | | | | | | 1 | | | | | | 1 |
| 40 | d[ɗ]=1 | | | | | 1 | | | | | | | | | 1 |
| 41 | Statystyka wierszy: | | | | | | | | | | | | | | |
| 42 | Suma głosek w wiersze: | 17 | 17 | 15 | 15 | 19 | 21 | 19 | 19 | 16 | 20 | 23 | 19 | 19 | |
| 43 | Suma głosek wyspich: | 1 | 1 | 2 | 0 | 2 | 2 | 0 | 2 | 1 | 2 | 1 | 1 | 1 | |
| 44 | Suma głosek średnich: | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 0 | 2 | 2 | 3 | 1 | 1 | |
| 45 | Suma głosek miękkich: | 6 | 6 | 3 | 4 | 4 | 4 | 3 | 6 | 4 | 4 | 2 | 5 | 5 | |
| 46 | Suma głosek twardych: | 2 | 2 | 2 | 3 | 2 | 2 | 4 | 0 | 3 | 2 | 3 | 3 | 1 | |
| 47 | Suma głosek twardej: | 7 | 7 | 4 | 5 | 9 | 11 | 7 | 10 | 5 | 10 | 11 | 9 | 9 | |
| 48 | Statystyka wiersza: | | | | | | | | | | | | | | |
| 49 | Suma głosek 239: | | | | | | | | | | | | | | |
| 50 | Suma głosek wyspich: | 16 | 6% | | | | | | | | | | | | |
| 51 | Suma głosek średnich: | 21 | 8% | | | | | | | | | | | | |
| 52 | Suma głosek miękkich: | 56 | 23% | | | | | | | | | | | | |
| 53 | Suma głosek twardych: | 29 | 12% | | | | | | | | | | | | |
| 54 | Suma głosek twardej: | 104 | 43% | | | | | | | | | | | | |
| 55 | | | | | | | | | | | | | | | |
| 56 | | | | | | | | | | | | | | | |

Rys. 10. Plik .csv z wynikami analizy statystycznej utworu

Źródło: opracowanie własne.

Dodatkowo, ponieważ autorzy utworów poetyckich często eksperymentują z rytmem w swoich utworach, a rytm potrafi zaburzyć schemat akcentuacji naturalny dla języka polskiego, użytkownik programu zawsze ma dostęp do zakładki „Błędy i ostrzeżenia”, w której można prześledzić komunikaty generowane podczas pracy narzędzia dla wszystkich akcji wykonywanych w czasie analizy utworu.

Wnioski końcowe i wskazania kierunków dalszych prac

W artykule zaprezentowano narzędzie komputerowe, które ma ułatwić badaczom literatury, w szczególności poezji, analizę wersyfikacji utworu poetyckiego oraz przyspieszyć proces anotacji wersologicznej utworów. Opisane narzędzie ma przyczynić się do rozwoju badań nad wersologią polskich utworów oraz pomóc w przeprowadzeniu badań zapoczątkowanych przez Marię Dłuską, dotyczących ikoniczności dźwięków mowy, szczególnie w oparciu o materiał poetycki.

Badania empiryczne i statystyczne mogą być (a może nawet: powinny być?), integralną częścią wszelkich działań interpretacyjnych utworu i dostarczają argumentów na rzecz subiektywnych odczuć czytelnika. Oczywiście, same liczby i statystyki nie wystarczą – należy pamiętać, że bez interpretacji wyników, a więc pracy badacza, liczby nic nie znaczą. Należy również pamiętać, że całkowite zaufanie metodom komputerowym może zaprowadzić badacza na manowce analityczne – wszelkie metody statystyczne działają najlepiej, gdy badania są przeprowadzane na odpowiednio dużych próbkach, należy również dbać o to, by interpretowane wyniki były istotne statystycznie. Wszelkie narzędzia komputerowe – które mają pomagać ludziom wykonywać pracę, która byłaby dla nich niewykonalna lub bardzo czasochłonna, gdyby nie oferowana przez oprogramowania szybkość i dokładność – są jedynie formą asysty, która dostarcza danych do analizy. Rola badacza jest tutaj więc tym większa, że to na nim spoczywa odpowiedzialność za czuwanie nad poprawnością badań. Komputer nie dostarczy nam gotowych odpowiedzi; nie można też bezwzględnie ufać wynikom, które przekazuje po przeprowadzeniu swojego procesu „myślowego”. Mając to na uwadze, można (i w obliczu wyraźnego postępu technologicznego w naukach – trzeba) integrować metody komputerowe w badaniach literackich. Mimo pewnych zagrożeń, które należy mieć w pamięci, narzędzia komputerowe oferują literaturoznawcom szereg rozwiązań, które z pewnością pomogą w uzupełnieniu interpretacji utworów o liczbowe analizy. SylLab może być jednak przyczynkiem do zwiększenia zainteresowania metodami obliczeniowymi w polskich badaniach literackich.

Tekst nie jest dowodem, że w każdym przypadku komputeryzacja językowej metody badawczej się powiedzie. Odwrotnie – artykuł pokazuje, że wykorzystanie komputera w tym zakresie często wiąże się z napisaniem autorskiego programu, który przynajmniej w części ułatwia lub porządkuje kolejne kroki

analizy. Podobna dyskusja na gruncie połączenia analizy literatury z matematyką toczy się już od dość dawna, a zwolenników takiego podejścia jest przypuszczalnie tyle samo, co jego przeciwników. W polskiej tradycji literaturoznawczej bardzo łatwo można doszukać się prób integracji tych dwóch – tylko pozornie – sprzecznych podejść do literatury. Szczególnie bogatą literaturę podejmującą ten temat odnaleźć można wśród prac publikowanych w czasach tzw. przełomu strukturalnego w Polsce. Należy tu wspomnieć m.in. o pracy pod redakcją Renaty Mayenowej *Poetyka i matematyka*¹¹ czy artykule Janusza Sławińskiego zatytułowanym *Wokół teorii języka poetyckiego*¹². Mimo dużej liczby prac wydawanych również przez Marię Dłuską, Lucyllę Pszczołowską czy Teresę Dobrzyńską w latach 50. i 60. XX wieku, niestety mało który współczesny interpretator – mimo zapewne dużego uznania wobec pracy wykonanej przez wspomnianych badaczy, nie sięga po zaproponowane przez nich metody badawcze. I chociaż temat integracji badań statystycznych w badaniach literackich nie jest tematem nowym, to może właśnie okres rozwoju technologicznego, który „odciąża” interpretatora i odejmuje mu obowiązek przeprowadzania analizy własnoręcznie, jest okresem, kiedy uda się tej integracji dokonać.

Artykuł nie jest tylko opisem konkretnego badania, które zresztą zostało oryginalnie przeprowadzone w sposób „manualny” wiele lat temu, a którego wyniki są powszechnie znane. Tekst pokazuje jednak, że podobną analizę można przeprowadzić w sposób wymagający znacznie mniejszego nakładu pracy i czasu. Autorka tekstu ma w planach przeprowadzenie analizy *Pana Tadeusza* Adama Mickiewicza, który jest tekstem zapewniającym stosunkowo duże próbki materiału do badań, które charakteryzują się jednolitym schematem rytmicznym, co niewątpliwie ułatwi badania dotyczące ikoniczności dźwięków mowy i powiązanie tego problemu z nastrojem poszczególnych fragmentów tekstu.

Prezentowane narzędzie może zostać użyte do analizy dużej liczby utworów literackich oraz ich automatycznej anotacji wersologicznej, co w przyszłości może przyczynić się do utworzenia pierwszego, ogólnodostępnego korpusu polskiej poezji. Taki korpus z pewnością ułatwiłby badania poetyckie, szczególnie w perspektywie komparatystycznej, a zawarta w takim potencjalnym korpusie analiza wersyfikacyjna utworów mogłaby spowodować większe zainteresowanie strukturą dźwiękową i rytmiczną utworów, która nadal skrywa przed czytelnikami oraz badaczami wiele tajemnic, czekających na odkrycie.

¹¹ R.M. Mayenowa, *Poetyka i matematyka*, Państwowy Instytut Wydawniczy, Warszawa 1965.

¹² J. Sławiński, *Wokół teorii języka poetyckiego*, w: tegoż, *Prace wybrane*, t. II: *Dzieło – język – tradycja*, Kraków 1998.

Bibliografia

- De Saussure Ferdinand, *Kurs językoznawstwa ogólnego*, przeł. Krystyna Kasprzyk, Warszawa 2007.
- Dłuska Maria, Kazimierz Wierzyński «Gdzie nie posiej mnie», w: *Liryka polska. Interpretacje*, red. Jan Prokop, Janusz Sławiński, słowo/obraz terytoria, Gdańsk 2001, s. 293–319.
- Iwańska Ilona, *Jasność i ciemność, czyli tonacje krzyżkowe i bemolowe*, „Teoria muzyki” 2018, nr 12, s. 105–125.
- Körner Anita, Rummer Ralf, *Articulation contributes to valence sound symbolism*, „Journal of Experimental Psychology General” 2021, nr 151(5), s. 1107–1114.
- Liryka polska. Interpretacje*, red. Jan Prokop, Janusz Sławiński, słowo/obraz terytoria, Gdańsk 2001.
- Mayenowa Maria Renata, *Poetyka i matematyka*, Państwowy Instytut Wydawniczy, Warszawa 1965.
- Nuckolls Janis, *The case for sound symbolism*, „Annu. Rev. Anthropol.” 1999, nr 28, s. 225–252.
- Ramachandran Vilayanur, *Neuronauka o podstawach człowieczeństwa*, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 2012, s. 129.
- Šeja Artjoms, Plecháč Petr, Lassche Alie, *Semantics of European poetry is shaped by conservative forces: The relationship between poetic meter and meaning in accentual-syllabic verse*, „PLoS ONE” 2022, nr 17 s. 1–17.
- Sławiński Janusz, *Wokół teorii języka poetyckiego*, w: tegoż, *Prace wybrane*, t. II: *Dzieło – język – tradycja*, Kraków 1998.
- Svantesson Jan-Olof, *Sound symbolism: the role of word sound in meaning*, „WIREs Cogn Sci” 2017, nr 8, s. 1–12.
- Śledziński Daniel, *Tworzenie reguł dla programu dzielącego tekst w języku polskim na sylaby*, „Biuletyn Polskiego Towarzystwa Językoznawczego” 2016, t. 72, s. 151–161.
- Trunin Mikhail, *Towards the concept of semantic halo*, „Studia Metrica et Poetica” 2017, t. 4, nr 2, s. 41–66.
- Wiśniewski Marek, *Zarys fonetyki i fonologii współczesnego języka polskiego*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń 2007.

Streszczenie

Wersologia jest dziedziną o bogatej tradycji, jednak ostatnio niestety zapominaną w badaniach poetyckich. Program Syllab powstał, by wrócić do korzeni polskiej wersologii i połączyć tradycję z nowymi technologiami. Narzędzie jest oparte na sposobie analizy wiersza Kazimierza Wierzyńskiego *Gdzie nie posiej mnie...* wykonanej przez Marię Dłuską. Program automatycznie przeprowadza analizę wersologiczną i fonetyczną utworu. Narzędzie jest nie tylko szansą na poprawę stanu polskich badań wersologicznych, ale także może przyczynić się do analizy większej liczby wierszy, co mogłoby umożliwić opracowanie pierwszego korpusu polskiej poezji.

How and Why to Automate Versological Analysis in Polish Literary Studies?

Abstract

Versology is a research domain with a rich tradition, however, partly forgotten in recent years. To return to the sources of Polish versology and combine tradition with modern technologies, the SylLab program was created. It is based on Maria Dłuska's interpretation of Wierzyński's poem *Gdzie nie posieją mnie...* It is a tool that automatically conducts versological and phonetical analysis of a poem. Such a tool not only improves poetic analysis and interpretation but can also contribute to the versological annotation of a considerable number of poems, possibly to start to create a corpus of Polish poems?

Słowa kluczowe: humanistyka cyfrowa, wersologia, wersyfikacja, automatyczna wersyfikacja, automatyczne badanie nastroju

Keywords: digital humanities, versology, versification, automatic versification, automatic sentiment analysis

Aleksandra Rykowska – studiowała lingwistykę komputerową na Wydziale Neofilologii oraz filologię polską na specjalności logopedycznej Wydziału Filologii Polskiej i Klasycznej Uniwersytetu im. Adama Mickiewicza. Absolwentka kierunków lingwistyka na Wydziale Filologicznym oraz filologia polska nauczycielska na Wydziale Polonistyki Uniwersytetu Jagiellońskiego w Krakowie. Obecnie doktorantka w Szkole Doktorskiej Nauk Humanistycznych UJ, gdzie przygotowuje pracę na temat stylometrii literackiego tłumaczenia maszynowego. Jej zainteresowania naukowe obejmują integrację i wykorzystanie nowych metod językoznawstwa, w szczególności językoznawstwa komputerowego, w analizie tekstów literackich. Interesuje się także językoznawstwem korpusowym, stylometrią i ilościowymi metodami badań tekstów, szczególnie poetyckich i ich wersologii.