

# **Odwołanie do fonologii w procesie dzielenia na sylaby wyrazów ortograficznych języka polskiego**

## **Reference to phonology in the process of syllabification of Polish words written in orthographic form**

*Daniel Śledziński*

UNIwersytet imienia Adama Mickiewicza w Poznaniu

danielsl@poczta.onet.pl

### **Abstrakt**

W artykule omówiono wybrane zagadnienia związane z procesem automatycznego dzielenia na sylaby tekstów ortograficznych w języku polskim. Przy projektowaniu systemu sylabifikującego uwzględniono kilka istotnych czynników: zasady fonologiczne, odczucia subiektywne, strukturę morfologiczną wyrazów. Zasady fonologiczne – przede wszystkim zasada sonorności – odnosi się do fonetycznej definicji sylaby. W niniejszej publikacji skoncentrowano się na aspektach działania systemu związanych z odwoływaniem się do zasad fonologicznych. Te zasady odnoszą się do wewnątrzwyrazowych grup spółgłoskowych, które są wyznaczane na podstawie informacji o lokalizacji liter odpowiadających ośrodkom sylab. To zadanie jest nieco utrudnione faktem, że w języku polskim niektóre litery, które zazwyczaj oznaczają samogłoski, w określonym kontekście oznaczają spółgłoski. W artykule omówiono również metodę przesłaniania podziału wynikającego z fonologii. Można stosować reguły które odnoszą się do odczucia subiektywnego i podział z niego wynikający różni się od podziału wynikającego z fonologii.

### **Abstract**

This paper presents selected topics related to automatic syllabification in Polish. Several factors have been taken into consideration in designing a syllabification system for Polish – phonological principles, subjective feelings, and the morphological structure of words. The phonological principles,

primarily the sonority principle, refer to the phonetic definition of a syllable, but subjective feelings sometimes suggest another division. The main goal of this article is to describe aspects of the syllabification system that refer to phonological principles. These principles refer to consonant clusters located within words, hence a method of extracting such clusters is given. This method is based on marking letters in the orthographic form of words that represent syllable nuclei. This task is somewhat complicated by the fact that some Polish letters that usually stand for vowels also represent consonants in particular neighborhoods. A method of masking the division based on phonology is also given. It is possible to apply rules that reflect subjective feelings, and the resulting division may differ from the division based on phonology.

## **1. Wprowadzenie**

W artykule omówiono wybrane zagadnienia związane z procesem automatycznego dzielenia na sylaby tekstów ortograficznych w języku polskim. Proces ten jest realizowany przez program komputerowy rozwijany przez autora artykułu (Śledziński 2013). W procesie dzielenia na sylaby mogą być brane pod uwagę następujące czynniki:

- zasady fonologiczne,
- odczucia subiektywne,
- struktura morfologiczna wyrazu.

Czynniki te mogą być uwzględniane w różnym stopniu, poza tym nie są one od siebie niezależne. Zasady fonologiczne umożliwiają podział zbliżony do podziału sugerowanego przez fonetyczną definicję sylaby, jednak podział taki nie zawsze jest zgodny z odczuciem subiektywnym. Natomiast niektóre granice morfologiczne są często odczuwane jako granice sylab, jednak jednocześnie mogą być one sprzeczne z zasadami fonologicznymi. Podstawowym celem autora jest stworzenie systemu elastycznego – takiego w którym wymienione czynniki mogłyby być uwzględnione w dowolnym zakresie. Cel ten może być zrealizowany poprzez tworzenie reguł o odpowiedniej strukturze.

W niniejszej publikacji skoncentrowano się na aspektach działania systemu związanych z odwoływaniem się do zasad fonologicznych oraz do odczuć, które mogą sugerować podział niezgodny z tymi zasadami. Zadania te są realizowane w odniesieniu do wewnątrzwyrazowych grup spółgłoskowych, które są wyznaczone na podstawie informacji o lokalizacji liter odpowiadających ośrodkom sylab. Dlatego znaczna część artykułu została poświęcona omówieniu zasad

lingwistycznych oraz algorytmowi, który wyznacza litery ortograficzne oznaczające ośrodki sylab.

## **2. Wyznaczanie ośrodków sylab w wyrazach ortograficznych**

### **2.1 Podstawy lingwistyczne**

Wyznaczanie ośrodków sylab oparto na informacjach zawartych w publikacji Marii Steffen-Batogowej: *Automatyzacja transkrypcji fonematycznej tekstów polskich* (Steffen-Batogowa 1975). System ortograficzny języka polskiego obejmuje 9 znaków, które mogą oznaczać samogłoski: *a, ą, e, ę, o, u, ó, y, i*. Można przyjąć, że litery: *a, e, o, ó, y* zawsze oznaczają ośrodek sylaby. Litery *ą* oraz *ę* oznaczają sekwens dwóch fonemów, którego struktura jest uzależniona od kontekstu. W zależności od przyjętej wymowy istotny może być tylko kontekst następujący lub zarówno kontekst poprzedzający, jak i kontekst następujący. Można uznać, że pierwszy człon sekwensu fonetycznego oznaczanego przez literę *ą* lub *ę* zawsze oznacza ośrodek sylaby. Omawiana struktura fonologiczna jest oznaczana przez jeden znak ortograficzny, zatem na płaszczyźnie ortograficznej jest ona nierozdzielna. Na potrzeby budowy algorytmu dzielącego na sylaby można zatem przyjąć uproszczone założenie, że litera *ą* oraz litera *ę* zawsze oznaczają ośrodek sylaby – założenie to wynika z oczywistego faktu, że nie ma możliwości wstawienia granicy słab wewnątrz sekwensu oznaczanego przez litery *ę* oraz *ą*.

Literze *u* stojącej po dowolnej literze różnej od litery *a* oraz litery *e* odpowiada zawsze fonem /u/. Literze *u* stojącej po literze *e* odpowiada fonem /w/. M.S.-B. wskazuje na wyjątki od tej reguły:

- jeżeli słowo *eu* stanowi fragment jednego z następujących słów: *ieu* (np.: *nieubłagany, nieuchronny, nieuk*), *rzeu* (np.: *przeubogi, przeuroczy*), *eusz* (np.: *chudeusz, jubileuszowy*), *eus#* (np. *kaduceus, Zeus*), *eum#* (np.: *apogeum, liceum*),
- we wszystkich formach wyrazów: *Pantateuch, reunion, teurgia*.

Natomiast literze *u* stojącej po literze *a* odpowiada fonem /w/ z wyjątkiem sytuacji opisanych przez M.S.-B. Według autorki litera *u* po literze *a* oznacza fonem /u/ w następujących sytuacjach:

- jeżeli litera *u* jest częścią jednego z następujących słów: *pozau* (np. *pozauczelniany*), *prau* (np. *praugrofński*), *dnau* (np. *przednaukowy*), *unau* (np. *unaukować*), *enau* (np. *nienaukowy*), *ezau* (np. *niezauważalny*), *onau* (np. *popularnonaukowy*), *anau* (np. *pozanaukowy*), *ynau* (np. *antynaukowy*);
- wszystkie formy fleksyjne wyrazów rozpoczynających się od *#nau* (np.: *naubliżać*, *nauka*, *nausznik*), za wyjątkiem form fleksyjnych następujących wyrazów: *nauplius*, *nautilus*, *nautofon*, *nautologia*, *nautologiczny*, *nautyczny*, *nautyka*.
- fonem /u/ odpowiada literze *u* w wyrazach rozpoczynających się od: *zau* przed dowolną literą różną od litery *t* – reguła ta dotyczy między innymi następujących wyrazów: *zaufać*, *zaułek*, *zauroczyć*, ale nie obejmuje wyrazu *zautomatyzować*;
- fonem /u/ odpowiada literze *u* w formach wyrazów: *laurka*, *laurkowy*, *asauł*, *auł*, *esauł*, *saksauł*.

W dalszym ciągu rozdziału przedstawiono informacje dotyczące litery *i*, która w języku polskim może oznaczać fonem /i/, fonem /j/ lub samodzielnie nie oznaczać żadnego fonemu (w niektórych kontekstach litera *i* jest tylko znakiem miękkości). Według M.S.-B. litera *i* nie oznacza fonemu /i/ w następujących sytuacjach:

- literze *i* w słowach należących do zbioru  $\{p, b, f, w, m, t, d, h, l, r\}iA$  z reguły odpowiada fonem /j/ (np.: *bariera*, *miliard*, *tiara*). Wyjątek stanowią złożenia wyrazowe takie jak: *biandria*, *miliamper*, *poliester*, w których litera *i* oznacza fonem /i/.
- W słowach o budowie  $\{p, b, f, w, m, t, d, h, l, r\}ii\#$  (np.: *kopiï*, *komediï*, *monarchiï*) pierwszej literze *i* może odpowiadać fonem /j/ lub fonem /i/ - autorzy, na których powołuje się M.S.-B., dopuszczają obie wersje.
- literze *i* występującej we wszystkich wyrazach pochodnych od wyrazów: *ekspiacja*, *kwietyzm*, *klient*, *miastenia*, *patriarcha*, *patriota* odpowiada słowo fonematyczne /ij/
- w słowach należących do zbioru  $\{\#, a\}\{t, p\}riA$  (np. w występujących w wyrazach: *trio*, *triumf*, *priorytet*, *patriota*) literze *i* odpowiada słowo fonematyczne /ij/.

- w słowach należących do zbioru  $\{s,c,z\}iA$  literze *i* odpowiada najczęściej słowo puste (np.: *siano, ziemia*). Wyjątek od tej reguły stanowią wyrazy: *siał, welwiczia*.
- literze *i* występującej w słowach ze zbioru  $niA$  odpowiada słowo puste. Reguła ta nie obejmuje wyrazów obcych (oraz wyrazów od nich pochodnych), które w mianowniku liczby pojedynczej są zakończone jednym ze słów ortograficznych: *nia* (np.: *mania, linia*), *nior* (np.: *senior, junior*), *nion* (np.: *kanion, kompanion*), *nium* (np.: *milenium, aluminium*)

## 2.2 Omówienie algorytmu

Algorytm, który wyznacza litery odpowiadające ośrodkom sylab w wyrazach ortograficznych, opiera się na informacjach lingwistycznych omówionych w rozdziale 2.1. Algorytm ten oparty jest na zestawie reguł i sprawia że, każdej literze w wyrazie ortograficznym przypisywany jest symbol C (łac.: *cousonans*) lub symbol V (łac.: *vocalis*), które oznaczają spółgłoskę lub samogłoskę. Dzięki temu można wyznaczyć ciągi liter, które odpowiadają wewnątrzwyrazowym grupom spółgłoskowym. W odniesieniu do tych fragmentów stosowane są metody omówione w następnych rozdziałach niniejszej publikacji.

W pierwszej kolejności następuje wyznaczenie wszystkich możliwych ciągów liter należących do danego wyrazu ortograficznego (w kolejności od fragmentów najdłuższych do pojedynczych liter). Zatem dla przykładowego wyrazu *kalka* będą to następujące fragmenty: *kalka, kalk, alka, kal, alk, lka, ka, al, lk, ka, k, a, l, k, a*. Następnie każdy uzyskany fragment porównywany jest z zestawem reguł (kolejność porównywania jest taka sama). Pojedyncza reguła ma prostą postać – składa się z dwóch członów. Pierwszy człon jest ciągiem znaków ortograficznych, natomiast drugi człon reguły jest odpowiednim ciągiem znaków C oraz V (ich liczba musi być równa liczbie liter w ciągu liter ortograficznych). W tabeli pierwszej przedstawiono wszystkie reguły, które są niezbędne dla wyznaczania liter odpowiadających ośrodkom sylab w wyrazach ortograficznych języka polskiego.

Tabela 1 – zestaw reguł dla wyznaczania ośrodków sylab w wyrazach ortograficznych.

1	<i>eu&gt;VC</i>	62	<i>mia&gt;CCV</i>	123	<i>cie&gt;CCV</i>
2	<i>ieu&gt;CVV</i>	63	<i>mio&gt;CCV</i>	124	<i>cia&gt;CCV</i>

*Daniel Śledziński: Odwołanie do fonologii w procesie dzielenia na sylaby  
wyrazów ortograficznych języka polskiego*

3	<i>rzeu&gt;CVV</i>	64	<i>mió&gt;CCV</i>	125	<i>cio&gt;CCV</i>
4	<i>eusz&gt;VVCC</i>	65	<i>müu&gt;CCV</i>	126	<i>ció&gt;CCV</i>
5	<i>eus#&gt;VVC#</i>	66	<i>miq&gt;CCV</i>	127	<i>ciu&gt;CCV</i>
6	<i>eum#&gt;VVC#</i>	67	<i>mię&gt;CCV</i>	128	<i>ciq&gt;CCV</i>
7	<i>tateuch&gt;CVCVVCC</i>	68	<i>tie&gt;CCV</i>	129	<i>cię&gt;CCV</i>
8	<i>au&gt;VC</i>	69	<i>tia&gt;CCV</i>	130	<i>zie&gt;CCV</i>
9	<i>pozau&gt;CVCVV</i>	70	<i>tio&gt;CCV</i>	131	<i>zia&gt;CCV</i>
10	<i>rau&gt;CCVV</i>	71	<i>tió&gt;CCV</i>	132	<i>zio&gt;CCV</i>
11	<i>dnau&gt;CCVV</i>	72	<i>tüu&gt;CCV</i>	133	<i>zió&gt;CCV</i>
12	<i>unau&gt;VCVV</i>	73	<i>tiq&gt;CCV</i>	134	<i>ziu&gt;CCV</i>
13	<i>enau&gt;VCVV</i>	74	<i>tię&gt;CCV</i>	135	<i>ziq&gt;CCV</i>
14	<i>ezau&gt;VCVV</i>	75	<i>die&gt;CCV</i>	136	<i>zię&gt;CCV</i>
15	<i>onau&gt;VCVV</i>	76	<i>dia&gt;CCV</i>	137	<i>nie&gt;CCV</i>
16	<i>anau&gt;VCVV</i>	77	<i>dio&gt;CCV</i>	138	<i>nia&gt;CCV</i>
17	<i>ynau&gt;VCVV</i>	78	<i>dió&gt;CCV</i>	139	<i>nio&gt;CCV</i>
18	<i>#nau&gt;#CVV</i>	79	<i>düu&gt;CCV</i>	140	<i>nió&gt;CCV</i>
19	<i>nauplius&gt; CVCCCCVC</i>	80	<i>diq&gt;CCV</i>	141	<i>niu&gt;CCV</i>
20	<i>nautil&gt;CVCCVC</i>	81	<i>dię&gt;CCV</i>	142	<i>niq&gt;CCV</i>
21	<i>nautof&gt;CVCCVC</i>	82	<i>hie&gt;CCV</i>	143	<i>nię&gt;CCV</i>
22	<i>nautol&gt;CVCCVC</i>	83	<i>hia&gt;CCV</i>	144	<i>a&gt;V</i>
23	<i>nautycz&gt;CVCCVC</i>	84	<i>hio&gt;CCV</i>	145	<i>q&gt;V</i>
24	<i>nautyk&gt;CVCCVC</i>	85	<i>hió&gt;CCV</i>	146	<i>b&gt;C</i>
25	<i>zau&gt;CVV</i>	86	<i>hiu&gt;CCV</i>	147	<i>c&gt;C</i>
26	<i>zaut&gt;CVCC</i>	87	<i>hiq&gt;CCV</i>	148	<i>ć&gt;C</i>
27	<i>laurk&gt;CVVCC</i>	88	<i>hię&gt;CCV</i>	149	<i>d&gt;C</i>
28	<i>laurc&gt;CVVCC</i>	89	<i>lie&gt;CCV</i>	150	<i>e&gt;V</i>
29	<i>asaul&gt;VCVVC</i>	90	<i>lia&gt;CCV</i>	151	<i>ę&gt;V</i>
30	<i>aul&gt;VVC</i>	91	<i>lio&gt;CCV</i>	152	<i>f&gt;C</i>
31	<i>esaul&gt;VCVVC</i>	92	<i>lió&gt;CCV</i>	153	<i>g&gt;C</i>
32	<i>saksaul&gt;CVCCVVC</i>	93	<i>liu&gt;CCV</i>	154	<i>h&gt;C</i>
33	<i>pie&gt;CCV</i>	94	<i>liq&gt;CCV</i>	155	<i>i&gt;V</i>
34	<i>pia&gt;CCV</i>	95	<i>lię&gt;CCV</i>	156	<i>j&gt;C</i>
35	<i>pio&gt;CCV</i>	96	<i>rie&gt;CCV</i>	157	<i>k&gt;C</i>
36	<i>pió&gt;CCV</i>	97	<i>ria&gt;CCV</i>	158	<i>l&gt;C</i>
37	<i>piu&gt;CCV</i>	98	<i>rio&gt;CCV</i>	159	<i>ł&gt;C</i>
38	<i>piq&gt;CCV</i>	99	<i>rió&gt;CCV</i>	160	<i>m&gt;C</i>
39	<i>pię&gt;CCV</i>	100	<i>riu&gt;CCV</i>	161	<i>n&gt;C</i>
40	<i>bie&gt;CCV</i>	101	<i>riq&gt;CCV</i>	162	<i>o&gt;V</i>
41	<i>bia&gt;CCV</i>	102	<i>rię&gt;CCV</i>	163	<i>ó&gt;V</i>
42	<i>bio&gt;CCV</i>	103	<i>biandr&gt;CVVCCC</i>	164	<i>p&gt;C</i>
43	<i>bió&gt;CCV</i>	104	<i>poliest&gt;CVCVVCC</i>	165	<i>r&gt;C</i>
44	<i>biu&gt;CCV</i>	105	<i>miliamper&gt; CVCVVCCVC</i>	166	<i>s&gt;C</i>
45	<i>biq&gt;CCV</i>	106	<i>püü#&gt;CCV#</i>	167	<i>ś&gt;C</i>
46	<i>bię&gt;CCV</i>	107	<i>büü#&gt;CCV#</i>	168	<i>t&gt;C</i>
47	<i>fie&gt;CCV</i>	108	<i>füü#&gt;CCV#</i>	169	<i>w&gt;C</i>
48	<i>fia&gt;CCV</i>	109	<i>wüü#&gt;CCV#</i>	170	<i>u&gt;V</i>

49	<i>fio</i> >CCV	110	<i>mü#</i> >CCV#	171	<i>y</i> >V
50	<i>fió</i> >CCV	111	<i>tü#</i> >CCV#	172	<i>z</i> >C
51	<i>fiu</i> >CCV	112	<i>dü#</i> >CCV#	173	<i>ź</i> >C
52	<i>fiq</i> >CCV	113	<i>hü#</i> >CCV#	174	<i>ż</i> >C
53	<i>fię</i> >CCV	114	<i>lü#</i> >CCV#		
54	<i>wie</i> >CCV	115	<i>rü#</i> >CCV#		
55	<i>wia</i> >CCV	116	<i>sie</i> >CCV		
56	<i>wio</i> >CCV	117	<i>sia</i> >CCV		
57	<i>wió</i> >CCV	118	<i>sio</i> >CCV		
58	<i>wiu</i> >CCV	119	<i>sió</i> >CCV		
59	<i>wiq</i> >CCV	120	<i>siu</i> >CCV		
60	<i>wię</i> >CCV	121	<i>siq</i> >CCV		
61	<i>mie</i> >CCV	122	<i>się</i> >CCV		

Ogólna zasada związana ze stosowaniem wymienionych w tabeli reguł jest taka, że reguły dłuższe (w sensie liczby znaków ortograficznych) mają większą moc. Zatem jeżeli w odniesieniu do pewnego fragmentu wyrazu została już zastosowana reguła dłuższa, to reguły krótsze (np. dotyczące pojedynczych liter) w odniesieniu do tego fragmentu nie będą stosowane. W tabeli pierwszej reguły o numerach 1 do 32 dotyczą przekształcania litery *u* po literze *e* oraz po literze *o*. Wiersze o numerach 33 do 143 dotyczą litery *i*. Natomiast pozostałe wiersze w tabeli pierwszej dotyczą pojedynczych znaków ortograficznych.

### 3. Wyznaczanie granic na podstawie zasad fonologicznych

#### 3.1 Podstawy lingwistyczne

Wyniki analizy dotyczącej wyznaczania granic sylab na podstawie zasad fonologicznych zostały omówione we wcześniejszych publikacjach autora. Również wcześniej przytaczano definicje fonologicznej zasady sonorności oraz fonologicznej zasady maksymalnego nagłosu<sup>1</sup> (Szypra-Kozłowska 2002). Stosowanie zasad fonologicznych można traktować jako uzupełnienie dla lingwistycznych definicji sylaby (Polański 1999), (Trask 1996), które są niewystarczające dla wyznaczania granic sylab – w szczególności wyrazów języka polskiego, w którym występuje znaczna liczba wieloelementowych grup spółgłoskowych o złożonej strukturze (Kuryłowicz 1952), (Dunaj 1985), (Dobrogowska 1984), (Dobrogowska 1990).

<sup>1</sup> Fonologiczna zasada sonorności nie dopuszcza spadku wartości sonorności dla kolejnych dźwięków w ramach nagłosu sylaby a także nie dopuszcza wzrostu wartości sonorności kolejnych dźwięków w ramach wygłosu sylaby. Na przykład w wyrazie /konto/, zgodnie z tą zasadą, zbitka /nt/ nie mogłaby należeć do nagłosu sylaby. Zasada sonorności umożliwia przydzielenie tej zbitki do wygłosu sylaby, jednak zasada maksymalnego nagłosu powoduje przydzielenie fonemu /t/ do nagłosu kolejnej sylaby, a zatem rozbitcie tej grupy spółgłoskowej między dwoma sylabami. We wcześniejszych badaniach przyjęto dodatkowe założenie, zgodnie z którym w ramach nagłosu oraz wygłosu sylaby nie była dopuszczalna równa sonorność kolejnych dźwięków.

Definicje fonologiczne sylaby opisują strukturę sylaby przy użyciu abstrakcyjnych klas dźwięków – samogłosek (ośrodków sylab) oraz spółgłosek (należących do marginaliów sylaby). Definicje fonetyczne odnoszą się do zjawisk fizycznych zachodzących w czasie artykulacji sylab – przede wszystkim do zmieniającego się stopnia rozwarcia narządów artykulacyjnych w czasie artykułowania sylab, przy czym na ośrodki sylab przypada maksymalne rozwarcie narządów artykulacyjnych, a na granice sylab przypada ich maksymalne zwarcie (Wierzchowska 1971). Zasady fonologiczne w abstrakcyjny sposób odzwierciedlają procesy fizyczne zachodzące w czasie artykulacji sylab. Fonologiczna skala sonorności odzwierciedla stopień rozwarcia narządów artykulacyjnych oraz donośność właściwą dla poszczególnych dźwięków mowy. Zatem spadek oraz wzrost fonologicznej wartości sonorności dla kolejnych fonemów w ramach grupy spółgłoskowej ułożonej między kolejnymi ośrodkami sylab odzwierciedla zmiany stopnia rozwarcia narządów artykulacyjnych oraz zmiany donośności kolejnych dźwięków mowy. Jednak trzeba podkreślić, że fonologia w znacznym stopniu upraszcza rzeczywistość fonetyczną – przypisywanie wartości sonorności do kolejnych dźwięków nie odzwierciedla wszystkich aspektów fonetycznych związanych z artykulacją tych dźwięków. Również założenie, że granica sylab musi być umiejscowiona między fonemami, nie musi mieć pełnego odzwierciedlenia w rzeczywistości – na przykład moment maksymalnego zwarcia może przypadać na środek artykulacji głoski (np. spółgłoski szczelinowej). Z drugiej jednak strony przytoczone już badania autora pokazały, iż przy założeniu, że w wygłosie sylaby wartość fonologicznej sonorności musi spadać, a w ramach nagłosu wartość ta musi wzrastać, można w sposób jednoznaczny wyznaczyć granice sylab dla ponad 99,5% wyrazów w tekstach korpusu. Podany wynik dotyczy przyjętej w tamtym badaniu skali sonorności (skala została zaproponowana przez Jolanę Szpyrę-Kozłowską). W jednej ze swoich publikacji autorka omawia również warianty skali sonorności zaproponowane przez innych badaczy (Szpyra-Kozłowska 1998). Swoje wersje przedstawili: Jespersen (1904), Hooper (1976), Selkirk (1984), Goldsmith (1990), Clements (1990, 1992).

Reasumując, fonologiczna zasada sonorności oraz fonologiczna zasada maksymalnego nagłosu stanowią skuteczny środek dla wyznaczania granic sylab. Zasady te w sposób abstrakcyjny i uproszczony odnoszą się do rzeczywistości fonetycznej, zatem w pewnym stopniu odzwierciedlają one treść fonetycznej



definicji sylaby. Stosowanie tych zasad pozwala na uzyskanie pewnej jednoznaczności. Czynnikiem zmiennym może być odniesienie do różnych skal sonorności. Granice uzyskane na podstawie zasad fonologicznych czasami są inne niż granice sugerowane przez strukturę morfologiczną wyrazu (problem ten był poruszony w przytoczonej już publikacji autora).

### **3.2 Działanie algorytmu**

Stosowanie zasad fonologicznych w odniesieniu do wyrazów ortograficznych jest kłopotliwe, ponieważ zasady te odnoszą się do wartości sonorności przypisywanej do dźwięków mowy (a nie do znaków ortograficznych). W języku polskim występują znaczne różnice ilościowe i jakościowe między płaszczyzną ortograficzną i płaszczyzną fonologiczną (i fonetyczną) (Ostaszewska, Tambor 2002), (Dunaj 2006). Jedna litera może oznaczać jeden lub dwa dźwięki mowy. Dwa lub trzy znaki ortograficzne mogą oznaczać jeden dźwięk mowy. Poza tym określony znak czy dwuznak może oznaczać różne fonemy (w zależności od kontekstu). W celu umożliwienia stosowania zasad fonologicznych w odniesieniu do wyrazów ortograficznych autor opracował system projekcji wewnątrzwyrazowych ortograficznych grup spółgłoskowych na transkrypcję fonologiczną. System ten zostanie omówiony dokładnie w oddzielnej publikacji, tutaj zostaną przedstawione jedynie informacje podstawowe. Omawiane rozwiązanie obejmuje wszystkie wewnątrzwyrazowe ortograficzne grupy spółgłoskowe języka polskiego. Każdej grupie ortograficznej odpowiada zapis tej samej grupy w transkrypcji fonologicznej Sampa. System przechowuje informacje o tym, które litery w zapisie ortograficznym odpowiadają poszczególnym fonemom w transkrypcji Sampa. System zawiera również niezbędne informacje o kontekście, w którym dana projekcja jest dopuszczalna. Dzięki projekcji można zastosować zasady fonologiczne w odniesieniu do ortograficznych grup spółgłoskowych w następujący sposób:

- najpierw w wyrazie ortograficznym są wyznaczone litery oznaczające ośrodki sylab – zgodnie z informacjami przedstawionymi w rozdziale drugim;
- na podstawie informacji o lokalizacji ośrodków sylab są wyznaczone ciągi liter będące wewnątrzwyrazowymi grupami spółgłoskowymi;

- wewnątrzwyrazowe ortograficzne grupy spółgłoskowe wraz z informacją o kontekście są przekazywane są do modułu projekcji, który zwraca transkrypcję fonologiczną danej grupy wraz z informacją o projekcji poszczególnych fonemów na zapis ortograficzny;
- przy odwołaniu do zasad fonologicznych następuje określenie miejsca podziału w zapisie transkrypcyjnym danej grupy;
- dzięki projekcji lokalizacja granicy zostaje przeniesiona na postać ortograficzną grupy i w tym miejscu zostaje wstawiony znacznik podziału, następnie dana grupa w wyrazie zostaje zamieniona wersją tej grupy ze wstawionym znacznikiem podziału.

Z wymienionych etapów szerszego omówienia wymaga etap dotyczący odwołania do zasad fonologicznych w celu określenia miejsca podziału w zapisie transkrypcyjnym danej grupy. Z informacji przedstawionych w rozdziale 3.1 wynika, że zgodnie z fonologią, granica powinna być wyznaczona w takim miejscu, aby wartość fonologicznej sonorności rosła w obu kierunkach w miarę zbliżania się do obu ośrodków sylab. Aby program mógł wykonać to zadanie, musi mieć informacje o wartości sonorności dla poszczególnych fonemów – wynika to z użytej skali sonorności. Algorytm analizuje wartość sonorności kolejnych fonemów i przed miejscem, w którym wartość ta zaczyna wzrastać, zostaje wstawiona granica sylab.

## **4. Odczucie subiektywne**

### **4.1 Wyniki badań**

We wstępie wspomniano, że podział oparty na zasadach fonologicznych może różnić się od podziału sugerowanego przez odczucia subiektywne. Rozważania te dotyczą odczuć subiektywnych, które nie wynikają ze struktury morfologicznej wyrazu, a jedynie z samej konstrukcji wewnątrzwyrazowej grupy spółgłoskowej. Autor przeprowadził dokładne badania dotyczące odczuć związanych z dzieleniem wyrazów ortograficznych na sylaby, przy czym były to wyrazy, których struktura morfologiczna nie sugerowała konkretnego rozwiązania. Dokładne wyniki tych badań zostaną przedstawione w kilku oddzielnych publikacjach. Tutaj została przywołana jedynie część ogólnych wyników w celu zasygnalizowania problemu oraz przedstawienia możliwości uwzględniania tego problemu w procesie dzielenia na sylaby.

W tabeli drugiej przedstawiono wyniki badania odczucia subiektywnego związanego z podziałem na sylaby wyrazów ortograficznych. Prezentację wyników zawężono do dwuelementowych (w sensie fonologicznym) wewnątrzwyrazowych grup spółgłoskowych. W badaniu udział wzięło 18 osób. Kolumna druga oraz kolumna trzecia wskazują na strukturę grupy (każdy wiersz tabeli dotyczy wszystkich wyrazów uwzględnionych w badaniu i zawierających grupę o danej strukturze). Natomiast kolumny oznaczone symbolami M1 oraz M2 zawierają informację o średniej liczbie odpowiedzi dla pierwszej oraz dla drugiej możliwości podziału. Pierwsza możliwość podziału wewnątrzwyrazowej dwuelementowej grupy spółgłoskowej dotyczy umiejscowienia granicy sylab przed pierwszą spółgłoską grupy. Druga możliwość podziału dotyczy wstawienia granicy przed drugą spółgłoską. Gwiazdką oznaczono możliwość podziału, która jest zgodna z podziałem wynikającym z zasad fonologicznych. Analizując te ogólne wyniki, łatwo zauważyć, że granica wynikająca z zasad fonologicznych nie zawsze pokrywa się z możliwością podziału preferowaną przez osoby biorące udział w eksperymencie. Rozbieżność między tymi dwoma czynnikami można wyjaśnić faktem, że zasady fonologiczne w ujęciu przedstawionym w rozdziale trzecim odnoszą się zjawisk fonetycznych, głównie do zmieniającego się stopnia rozwarcia narządów artykulacyjnych w czasie artykułowania sylab. Natomiast wpływ na odczucia subiektywne mogą mieć również własności fonotaktyczne oraz własności fonostatystyczne związane z poszczególnymi grupami spółgłoskowymi. (Jassem, Łobacz 1974). Te zagadnienia zostaną szerzej omówione w oddzielnych publikacjach poświęconych odczuciu subiektywnemu związanemu z podziałem na sylaby.

Tabela 2 – wyniki badania dotyczące odczucia subiektywnego związanego z dzieleniem na sylaby.

Lp.	I spółgł.	II spółgł.	M1	M2
1	Zw	Zw	2.1	*15.7
2	Zw	Sz	*5.2	12.6
3	Sz	Zw	2.5	*15.4
4	Sz	Sz	3.0	*15.0
5	Zw	Pó	*9.7	8.2
6	Zw	Pł	*11.2	6.4
7	Zw	N	*5.8	12.0

8	Sz	Pó	*11.3	6.7
9	Sz	Pł	*7.6	10.4
10	Sz	N	*4.0	13.9
11	Pó	Zw	1.0	*16.9
12	Pó	Sz	1.0	*16.9
13	Pó	Pó	2.0	*16.0
14	Pó	Pł	1.5	*16.5
15	Pó	N	0.8	*17.0
16	Pł	Zw	0.9	*17.1
17	Pł	Sz	1.1	*16.6
18	Pł	Pó	*1.7	16.0
19	Pł	Pł	1.3	*16.3
20	Pł	N	1.2	*16.8
21	N	Zw	0.8	*17.2
22	N	Sz	1.2	*16.8
23	N	Pó	*4.5	13.5
24	N	Pł	*2.3	15.5
25	N	N	2.3	*15.5
26	Zw	Zw	2.1	*15.7
27	Zw	Sz	*5.2	12.6
28	Sz	Zw	2.5	*15.4
29	Sz	Sz	3.0	*15.0

## 4.2 Działanie algorytmu

Rozbieżności pomiędzy podziałem wynikającym z fonologii oraz z wyników badania odczuć subiektywnych można uwzględnić w procesie podziału. W tym celu należy stworzyć dodatkowe reguły dla określonego zbioru grup spółgłoskowych, które będą przesłaniać podział wynikający z fonologii. Aby mechanizm przesłaniania mógł funkcjonować skutecznie należałoby zmodyfikować procedurę omówioną w rozdziale 3.2. Przed krokiem, w którym następuje odwołanie do zasad fonologicznych, musi zostać sprawdzone, czy istnieje dodatkowa reguła dla danej grupy spółgłoskowej. Jeżeli dodatkowy zbiór reguł zawiera taką regułę, to zostanie ona zastosowana, natomiast krok obejmujący odwołanie do zasad fonologicznych zostanie pominięty. W ten sposób można zapewnić podział również dla tych grup spółgłoskowych, których nie da się podzielić przy odwołaniu do zasad fonologicznych – na przykład grup, w ramach których następuje wzrost oraz spadek wartości sonorności (np. w wyrazie *pasłby*).

## 5. Podsumowanie

W artykule przedstawiono szczegółowe informacje dotyczące stosowania zasad fonologicznych w rozwijanym systemie przeznaczonym do dzielenia na sylaby

wyrazów ortograficznych w języku polskim. Podstawowym celem zaprezentowanych rozwiązań jest odejście od podejścia statycznego opartego na zamkniętym zbiorze reguł. Celem przedstawionych rozwiązań jest umożliwienie odniesienia do różnych czynników mogących mieć wpływ na podział na sylaby oraz umożliwienie regulowania stopnia odniesienia do tych czynników. Dzięki takiemu podejściu można stosować na przykład skale sonorności różnych autorów oraz uwzględnić podział zgodny z subiektywnym odczuciem użytkowników języka.

W artykule omówiono kilka istotnych lingwistycznych zagadnień. Przede wszystkim przedstawiono algorytm oraz zbiór reguł służących do wyznaczania ośrodków sylab w wyrazach ortograficznych. Omówiono również inne etapy związane z odniesieniem do zasad fonologicznych, między innymi etap polegający na odwołaniu do modułu projekcji zapisu ortograficznego wewnątrzwyrazowych grup spółgłoskowych na transkrypcję fonologiczną czy też opcjonalny etap, w którym następuje odwołanie do reguł modyfikujących podział wynikający z fonologii. Niektóre zagadnienia, o których wspomniano w tym artykule zostały lub będą szczegółowo omówiony w oddzielnych publikacjach. Dotyczy to między innymi wspomnianego już modułu projekcji zapisu ortograficznego grup spółgłoskowych czy też szczegółowych wyników badań odnoszących się do odczucia subiektywnego związanego z dzieleniem na sylaby wyrazów ortograficznych w języku polskim.

#### **Bibliografia:**

- Dobrogowska K. 1984. Śródgłosowe grupy spółgłosek w polskich tekstach popularnonaukowych. *Polonica*, X, s. 15-34.
- Dobrogowska K. 1990. Word internal consonant clusters in Polish artistic prose, *Studia Phonetica Posnaniensia*, II, s. 43-67.
- Dunaj B. 1985. Grupy spółgłoskowe współczesnej polszczyzny mówionej (w języku mieszkańców Krakowa). W: *Prace Językoznawcze, DCCCIII..* Nakł. Uniwersytetu Jagiellońskiego, PWN. Kraków, Warszawa. s. 46-79.
- Dunaj B. 2006. Zasady poprawnej wymowy polskiej. *Język Polski* 86, 3. s. 161–172.
- Jassem W., Łobacz P. 1974. Fonotaktyczna analiza mówionego tekstu polskiego. *Biuletyn Polskiego Towarzystwa Językoznawczego*, XXXII. s. 179-197.
- Kuryłowicz J. 1952. Uwagi o polskich grupach spółgłoskowych. *BPTJ* 11. s. 54–69.
- Ostaszewska D., Tambor J. 2002. *Fonetyka i fonologia współczesnego języka polskiego.* Wydawnictwo Naukowe PWN. Warszawa. s. 47-83.

*Daniel Śledziński: Odwołanie do fonologii w procesie dzielenia na sylaby  
wyrazów ortograficznych języka polskiego*

- Polański K. red. 1999. Encyklopedia językoznawstwa ogólnego. Zakład Narodowy im. Ossolińskich – Wydawnictwo. Wrocław. s. 644.
- Szpyra-Kozłowska J. 1998. The sonority scale and phonetic syllabification in Polish. *Biuletyn Polskiego Towarzystwa Językoznawczego*, LIV. s. 63-82.
- Szpyra-Kozłowska J. 2002. Wprowadzenie do współczesnej fonologii. Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej. Lublin. s. 150-152.
- Steffen-Batogowa M. 1975. Automatyzacja transkrypcji fonematycznej tekstów polskich. PWN. Warszawa. s. 78-90.
- Śledziński D. 2013. Podział korpusu tekstów na sylaby – analiza polskich grup spółgłoskowych. *Kwartalnik Językoznawczy*, XV. s. 48-100.
- Trask R. L. 1996. A dictionary of phonetics and phonology. Routledge. New York, s. 345.
- Wierzchowska B. 1971. Wymowa Polska. Państwowe Zakłady Wydawnictw Szkolnych, wyd. II. Warszawa. s. 214.