

DANIEL ŚLEDZIŃSKI

Uniwersytet im. Adama Mickiewicza w Poznaniu  
Poznań

## **Tworzenie reguł dla programu dzielącego tekst w języku polskim na sylaby**

### **Streszczenie**

Dostępne lingwistyczne definicje sylaby nie umożliwiają jednoznacznego podziału wyrazów na sylaby. Dotyczy to w szczególności języka polskiego, w którym występują liczne wieloelementowe grupy spółgłoskowe niespotykane w innych językach. Niniejszy artykuł dotyczy programu komputerowego (rozwijanego przez autora) przeznaczonego dla dzielenia tekstów w języku polskim na sylaby. Działanie tego programu opiera się na wynikach wcześniejszych badań autora, które wykazały istotne znaczenie zarówno fonologii, jak i morfologii dla procesu dzielenia wyrazów na sylaby. Badania te wykazały, że podział wyrazów w języku polskim prawie zawsze można oprzeć na zasadach fonologicznych, jednak w niektórych przypadkach morfologia sugeruje inny podział. Dlatego podstawowa procedura podziału oparta jest na regułach wynikających z fonologicznej zasady sonorności, natomiast stosowanie reguł wynikających ze struktury morfologicznej wyrazów jest opcjonalne. Artykuł dotyczy rozwiązań zastosowanych w programie – przede wszystkim omówiona została struktura reguł podziału. Publikacja porusza również problem różnic między zapisem ortograficznym i transkrypcją ortograficzną w języku polskim. Ten problem jest istotny, ponieważ proces dzielenia na sylaby dotyczy tekstów ortograficznych, natomiast wartości fonologicznej sonorności są przypisywane do dźwięków mowy.

**Słowa kluczowe:** sylaba, podział na sylaby, fonologia, morfologia, zasada sonorności, program komputerowy, reguły.

## **Designing rules for a computer program purposed for Polish texts syllabification**

### **Abstract**

According to the current definitions of a syllable it is impossible to syllabify words in an unambiguous and uncontroversial manner. This especially concerns languages such as Polish, which is rich in long consonant clusters, rarely occurring in other languages. The article describes computer software cur-

rently being developed by the author to divide Polish texts into syllables. The program is based on the findings of the author's previous research, which point to the significance of both phonology and morphology for syllabification in Polish. Investigation has shown that syllabification may, in most cases, be based mostly on phonology, but sometimes morphology suggests different solutions. Therefore the main syllabification procedure is based on the phonological principle of sonority, with an optional syllabification procedure based on the morphological structure of words. The article outlines the solutions employed in the program – mainly the structure of syllabification rules. Attention is also drawn to the differences between orthographic spelling and phonological transcription in Polish, which is important due to the fact that the syllabification program operates on orthographic words, while phonological sonority values are applicable to speech sounds.

**Key words:** syllable, syllabification, phonology, morphology, sonority sequencing principle, computer program, rules.

## 1. Wprowadzenie

W artykule omówiono program komputerowy przeznaczony do dzielenia tekstu w języku polskim na sylaby. Program jest rozwijany przez autora artykułu<sup>1</sup>. Artykuł dotyczy przede wszystkim aspektów lingwistycznych związanych z tworzeniem reguł dla tego programu.

Dostępne lingwistyczne definicje sylaby nie umożliwiają jednoznacznego podziału wyrazów na sylaby. Definicje fonologiczne zakładają występowanie obligatoryjnego ośrodka sylaby (samogłoski) oraz fakultatywnych marginaliów – nagłosu oraz wygłosu sylaby (złożonych ze spółgłoski lub z grupy spółgłoskowej) (Trask 1996: 345) (Polański 1999: 644). Zatem definicje te nie precyzują sposobu wyznaczania granic sylab – w szczególności dla wieloelementowych grup spółgłoskowych (ulożonych między kolejnymi ośrodkami sylab), które w języku polskim mają wyjątkowo złożoną strukturę (Dunaj 1985: 46–79, Dobrogowska 1984: 15–34, Dobrogowska 1990: 43–63, Jassem, Łobacz 1974: 179–197). Natomiast definicje fonetyczne odnoszą się do zjawisk fizycznych zachodzących w czasie artykułowania sylab – przede wszystkim do zmian dotyczących: stopnia rozwarcia narządów artykulacyjnych, ciśnienia subglotalnego i natężenia fali głosowej (Wierzchowska 1971: 214). Maksymalne rozwarcie narządów artykulacyjnych przypada na ośrodek sylaby, a ich maksymalne zwarcie przypada na krańce sylaby. Jednak taka definicja jest niewystarczająca dla jednoznacznego określania granic między wszystkimi sylabami, ponieważ określenie miejsce maksymalnego zwarcia narządów artykulacyjnych może być kłopotliwe (w szczególności dla wieloelementowych grup spółgłoskowych). Poza tym miejsce to może w rzeczywistości przypadać na środek artykulacji niektórych spółgłosek (np. spółgłosek szczelinowych).

---

<sup>1</sup> Program zostanie przekazany do domeny publicznej – będzie mógł być używany nieodpłatnie dla dowolnych celów. Planowane jest udostępnienie programu oraz różnych wersji reguł podziału po adresem: [www.sylabifikator.pl](http://www.sylabifikator.pl).

Wcześniejsze badania autora (Śledziński 2013: 48–100) wykazały kilka istotnych faktów, które dotyczą dzielenia wyrazów w języku polskim na sylaby. Przede wszystkim potwierdzona została przydatność fonologicznych zasad – zasady sonorności oraz zasady maksymalnego nagłosu. W badaniach uwzględniono też strukturę morfologiczną wyrazów – wykazano, że niektóre granice morfologiczne, które mogą być odczuwalne jako naturalne granice sylab, są sprzeczne z podziałem wynikającym z zasad fonologicznych. Te fakty zostały uwzględnione przy projektowaniu omawianego programu. Program umożliwi uwzględnienie fonologii i morfologii lub tylko fonologii w procesie dzielenia na sylaby. Poza tym stopień uwzględnienia tych czynników również można regulować – zagadnienia te zostały szczegółowo omówione w dalszym ciągu publikacji. Wszystkie rozwiązania opisane w artykule mają charakter umowy – są one przyjmowane dla celów praktycznych związanych z dzieleniem na sylaby tekstów w języku polskim.

## 2. Odniesienie do fonologii

### 2.1. Istotne definicje

Fonologiczna skala sonorności przypisuje poszczególnym fonemom abstrakcyjne wartości liczbowe odzwierciedlające właściwości fonetyczne dźwięków mowy – przede wszystkim stopień rozwarcia narządów artykulacyjnych. We wcześniejszych badaniach została użyta skala sonorności zaproponowana przez Jolantę Szpyrę-Kozłowską dla języka polskiego (Szpyra-Kozłowska 2002: 150). Skala ta została przytoczona w tabeli pierwszej. W innej publikacji autorka przytacza i porównuje skale sonorności zaproponowane przez różnych autorów (Szpyra-Kozłowska 1998: 63–82).

Klasa głosek	Sonorność
Samogłoski	6
Półsamogłoski	5
Spółgłoski płynne	4
Spółgłoski nosowe	3
Spółgłoski szczelinowe	2
Spółgłoski zwarte: zwarto-wybuchowe oraz zwarto-szczelinowe	1

Tabela 1. Skala sonorności przyjęta w badaniach

Zgodnie z fonologiczną zasadą sonorności – sonorność dźwięków powinna maleć w miarę oddalania się od ośrodka sylaby i rosnać w miarę zbliżania się do kolejnego ośrodka sylaby<sup>2</sup>. Z kolei zasada maksymalnego nagłosu zakłada przydzielanie do

<sup>2</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

nagłosu sylaby maksymalnej liczby spółgłosek – o ile nie jest to sprzeczne z zasadą sonorności (Szpyra-Kozłowska 2002: 151–152). Wspomniane w rozdziale pierwszym badania autora wykazały, że na podstawie tych zasad fonologicznych można wyznaczyć granice sylab dla 99,7% wyrazów w tekstach w języku polskim.

Przytoczone zasady fonologiczne w sposób abstrakcyjny odzwierciedlają procesy fizyczne zachodzące w czasie artykułowania sylab. Na podstawie ustaleń umownych można pominąć niektóre dyskusyjne zagadnienia – na przykład fakt, że moment maksymalnego zwarcia narządów artykulacyjnych niekoniecznie musi przypadać między głoskami – dla uproszczenia można przyjąć, że tak właśnie jest. Również dyskusji można poddawać zasadność przypisywania abstrakcyjnych wartości liczbowych do klas dźwięków mowy – taki zabieg nie odzwierciedla całej złożoności procesu artykulacyjnego. Zatem stosowanie tych zasad fonologicznych wiąże się z pewnym uproszczeniem rzeczywistości fonetycznej, jednak jest to potrzebne, aby zapewnić jednoznaczność i konsekwentność przy dzieleniu wyrazów na sylaby.

## 2.2. Reguły podziału

Podział wynikający z zasad fonologicznych dotyczy wewnątrzwyrazowych grup spółgłoskowych (lub wewnątrzwyrazowych pojedynczych spółgłosek). Dla omawianego programu trzeba stworzyć listę reguł, która jest zapisana w specjalnym pliku tekstowym. Jedna reguła obejmuje zapis ortograficzny wewnątrzwyrazowej grupy spółgłoskowej (umieszczonej w konkretnym kontekście samogłoskowym) oraz zapis ortograficzny tej samej grupy spółgłoskowej z wstawionym znacznikiem granicy sylab (pionowa kreska). Miejsce wstawienia znacznika podziału wynika z zasad fonologicznych, przyjętej skali sonorności, z dodatkowych założeń lub z innych czynników uwzględnionych przy budowie danej listy reguł podziału. Przykłady takich par ujęto w tabeli drugiej<sup>3</sup>.

Ciąg liter przed zamianą	Ciąg liter po zamianie	Przykładowy wyraz
elbi	el bi	wielbiciel
elbia	el bia	uwielbiać
elbią	el bią	niewielbiąca

Tabela 2. Przykłady reguł podziału opartych na zasadach fonologicznych

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.

<sup>3</sup> Taki sposób zapisu reguł przewiduje wyszczególnienie wszystkich kontekstów samogłoskowych. Aktualnie planowana jest również inna struktura reguł podziału dla wewnątrzwyrazowych grup spółgłoskowych. Przewiduje ona określanie kontekstu grup spółgłoskowych przy użyciu oznaczonych zbiorów samogłosek – taki zabieg znacznie zmniejszy liczbę reguł. Poza tym taki zapis może być bez przeszkód konwertowany do zapisu ujętego w tabeli 2.

Aby stosować takie reguły, trzeba zastosować algorytm, który wskazuje w wyrazach ortograficznych litery oznaczające ośrodki sylab – w celu wyznaczania fragmentów „od samogłoski do samogłoski” oraz podmieniania tych fragmentów odpowiednimi fragmentami zawierającymi znaczniki granic sylab. Algorytm ten stworzono na podstawie zasad wymowy i reguł transkrypcji omówionych przez Marię Steffen-Batogową w książce: *Automatyzacja transkrypcji fonematycznej tekstów polskich* (1975: 78–90). Na podstawie tej publikacji wyszczególniono następujące zasady dotyczące wyznaczania ośrodków sylab w wyrazach ortograficznych:

- jeżeli po literze *i* występuje samogłoska, to litera ta nigdy nie jest ośrodkiem sylaby, z wyjątkiem wyrazów rozpoczynających się od ciągów liter: *poli* lub *mili* (np.: *poliester*, *miliamper*). Natomiast jeżeli litera *i* występuje przed spółgłoską, to oznacza ona ośrodek sylaby (fonem /i/);
- jeżeli litera *u* występuje po dowolnej literze ortograficznej różnej od *a* oraz *e*, to zawsze oznacza ona fonem /u/ (jest ośrodkiem sylaby). Jeżeli litera *u* jest fragmentem któregoś ze słów ortograficznych: *ieu* (np. *nieubłaganay*), *rzeu* (np. *przeubogi*), *eusz* (np. *jubileuszowy*), *eus#* (np. *Zeus*), *eum* (np. *liceum*), *pozau* (np. *pozauczelniany*), *prau* (np. *praugrofiński*), *nauk* (np. *przednaukowy*), *#nau* (*naubliżać*, *nauczać*), to także oznacza ona zgłoskotwórczy fonem /u/. Również w wyrazach: *laurka*, *asaul* oraz w formach fleksyjnych tych wyrazów litera *u* oznacza fonem /u/. W pozostałych wyrazach litera *u* po literach *a* lub *e* oznacza niezgłoskotwórczy fonem /w/;
- litera *y* prawie zawsze oznacza fonem /y/ – istnieją nieliczne wyjątki (*Domeyko*, *yeti*),
- litery *q* oraz *ę* oznaczają sekwens samogłoski (/o/ lub /e/) oraz półsamogłoski lub spółgłoski nosowej. Zatem litery te zawsze oznaczają obecność ośrodka sylaby;
- litery *a*, *e*, *o*, *ó* zawsze oznaczają ośrodki sylab.

### 2.3. Projekcja zapisu ortograficznego na transkrypcję fonologiczną

Wartości sonorności można określać dla dźwięków mowy (lub dla klas dźwięków mowy). Zatem zasadę sonorności oraz zasadę maksymalnego nagłosu można stosować tylko dla tekstu transkrybowanego (w odniesieniu do fonemów, a nie do znaków ortograficznych). Jednak reguły podziału obejmujące wewnątrzwyrazowe grupy spółgłoskowe oraz otaczające te grupy samogłoski są zapisywane przy użyciu transkrypcji ortograficznej (por. rozdz. 2.2). Przy tworzeniu takich reguł można odwołać się do dowolnej skali sonorności oraz uwzględnić dowolne dodatkowe założenia. Jednak „ręczne” nanoszenie granic dla kilku tysięcy reguł podziału jest niewątpliwie procesem uciążliwym i czasochłonnym. Dlatego poczyniono starania zmierzające w kierunku automatyzacji tego procesu – rozwijany jest oddzielny moduł projekcji zapisu ortograficznego wewnątrzwyrazowych grup spółgłoskowych na transkrypcję fonologiczną. Zabieg ten polega na przypisaniu kolejnych znaków ortograficznych do odpowiednich fonemów. Dzięki temu można automatycznie przypisywać wartości fonologicznej sonorności do odpowiednich liter ortograficznych (lub par liter), a to

z kolei pozwala na automatyzację stosowania zasad fonologicznych w odniesieniu do ortograficznych grup spółgłoskowych należących do reguł podziału oraz na automatyzację wstawiania znaczników granic sylab w tych regułach. Tabela 3 oraz tabela 4 zawierają przykłady projekcji wewnątrzwyrazowych grup spółgłoskowych na odmianę transkrypcji fonologicznej SAMPA utworzonej dla języka polskiego (Demenko 2003: 79–95).

Indeks	0	1	2	3
Ortografia	o	t	ni	e
SAMPA	/o/	/t/	/n'/	/e/
Sonorność	6	1	3	6

Tabela 3 Projekcja łańcucha *otnie* na transkrypcję fonologiczną

Indeks	0	1	2	3
Ortografia	ę	b	sz	y
SAMPA	/em/	/p/	/S/	/y/
Sonorność	6-3	1	2	6

Tabela 4. Projekcja łańcucha *ębszy* na transkrypcję fonologiczną

Przedstawione w tabelach przykłady dotyczą konkretnych kontekstów samogłoskowych. Jednak w praktyce takie tabele można uogólnić (jednocześnie znacznie zmniejszając ich liczbę). Zamiast podawania konkretnego kontekstu samogłoskowego można określić dopuszczalne (potencjalne) zbiory samogłosek, które mogą otaczać projekcję danej grupy spółgłoskowej. Moduł projekcji zapisu ortograficznego wewnątrzwyrazowych grup spółgłoskowych na zapis fonologiczny jest aktualnie rozwijany i temu zagadnieniu zostaną poświęcone oddzielne publikacje. W niniejszym artykule warto zasygnalizować problemy związane z tym zagadnieniem:

- ilościowe i jakościowe różnice między płaszczyzną fonologiczną i płaszczyzną ortograficzną w języku polskim: jedna litera może oznaczać różne fonemy (w zależności od kontekstu), dwie litery mogą oznaczać jeden fonem (dwuznaki), redukcje przy wymowie niektórych grup spółgłoskowych (Ostaszewska, Tambor 2002: 47–83);
- różnice w inwentarzach fonologicznych różnych autorów – niektóre dźwięki mogą być uznawane za alofony określonych fonemów lub za oddzielne fonemy;
- uzależnienie wymowy niektórych wewnątrzwyrazowych grup spółgłoskowych od szerszego kontekstu (wykraczającego poza samogłoski otaczające tę grupę spółgłoskową). Liczba takich wyrazów nie jest duża – problem ten jest związany przede wszystkim z parami liter, które najczęściej są dwuznakami (oznaczają jeden fonem), jednak czasami mogą oznaczać dwa fonemy (np.: *marznę, odzew*).

### 3. Odniesienie do morfologii

#### 3.1. Istota problemu

Niektóre granice morfologiczne mogą być odczuwalne przez użytkowników języka jako naturalne granice sylab – w szczególności granice między przedrostkiem i rdzeniem wyrazu. Takie granice są często ulokowane w innym miejscu niż granice wynikające ze stosowania zasad fonologicznych. Autor przeprowadził testy odczuć subiektywnych związanych z dzieleniem wyrazów ortograficznych na sylaby. Okazało się, że dla takich problematycznych przypadków struktura morfologiczna (tzn. granica między przedrostkiem a rdzeniem) ma najczęściej większy wpływ na wyniki testów. Podobnie w *Wielkim słowniku ortograficznym PWN* przy omawianiu reguł związanych z dzieleniem wyrazów ortograficznych (przy przenoszeniu między wierszami) podkreślono przewagę czynnika morfologicznego nad czynnikiem fonetycznym (Połański 2006: 85–87).

Jeżeli rozważania ograniczyć tylko do granic między przedrostkiem i rdzeniem, to odczuwalna wyrazistość takich granic również może być zróżnicowana dla różnych wyrazów. Omawiane fakty uwzględniono przy projektowaniu programu – umożliwia on definiowanie listy reguł dotyczących dowolnych przedrostków i dowolnych wyrazów, przy czym reguły te mogą przesłaniać reguły oparte na zasadach fonologicznych (por. rozdz. 2).

Niezgodność podziału wynikającego z fonologii i z morfologii dotyczy następujących przypadków:

- jeżeli przedrostek zakończony jest samogłoską, a na początku rdzenia ulokowana jest grupa spółgłoskowa, w której wartość sonorności kolejnych elementów maleje (np.: *posprzątać*, *dostawić*);
- jeżeli przedrostek zakończony jest spółgłoską, a rdzeń rozpoczyna się od samogłoski (np.: *poduczyć się*);
- jeżeli przedrostek zakończony jest spółgłoską, a rdzeń rozpoczyna się od spółgłoski o sonorności większej niż spółgłoska należąca do przedrostka (np.: *podjechać*);
- poza wymienionymi przypadkami w wyniku połączenia spółgłoski lub grupy spółgłoskowej należącej do przedrostka ze spółgłoską lub grupą spółgłoskową należąca do rdzenia wyrazu może powstać grupa spółgłoskowa, której nie da się podzielić na podstawie zasad fonologicznych (np.: *odsprzedaje*). W takich przypadkach podział wynikający ze struktury morfologicznej nie ma alternatywy w podziale opartym na zasadach fonologicznych.

#### 3.2. Struktura i stosowanie reguł dla przedrostków

Można stworzyć dowolną liczbę reguł dla przedrostków. Istnieją pewne istotne utrudnienia, które muszą być uwzględnione w mechanizmie stosowania reguł oraz w samych regułach. Przede wszystkim język polski cechuje się złożoną fleksją – zatem

tworzone reguły muszą uwzględniać wszystkie formy fleksyjne poszczególnych form podstawowych wyrazów. Drugie istotne utrudnienie związane jest z faktem, że jeden ciąg znaków ortograficznych może jednocześnie zawierać różne przedrostki i być częścią różnych wyrazów. W dalszym ciągu niniejszego rozdziału omówiono przykłady, które obrazują mechanizm działania reguł, ich strukturę oraz ukazują rozwiązania wymienionych problemów.

W przeciwieństwie do reguł omówionych w rozdziale drugim reguły dotyczące przedrostków wykraczają poza wewnątrzwyrazowe grupy spółgłoskowe oraz otaczające je samogłoski. Reguły te muszą uwzględniać dany przedrostek oraz kolejne litery, na podstawie których można zidentyfikować jak największą liczbę form fleksyjnych danego wyrazu. Zamiast wyznaczania łańcuchów znaków „od samogłoski do samogłoski” algorytm najpierw wyznacza wszystkie możliwe łańcuchy liter rozpoczynające się od pierwszej litery danego wyrazu. Każdy z otrzymanych w ten sposób łańcuchów porównywany jest z dostępnymi regułami podziału – w celu ewentualnej zamiany na taki sam łańcuch zawierający znacznik granicy sylab. Na przykład dla wyrazu *podbarwiłem* zostaną wyznaczone następujące łańcuchy znaków:

*podbarwiłem, podbarwiłe, podbarwił, podbarwi, podbarw, podbar, podba, podb, pod, po, p.*

Każdy z tych fragmentów zostanie porównany z zestawem reguł dotyczących przedrostków. Zestaw ten musi zawierać regułę (lub reguły), która zapewni podział wszystkich form fleksyjnych formy podstawowej: *podbarwić*. Taką regułę można zbudować na podstawie ciągu znaków: *podbarw*. Zatem reguła: *podbarw – pod|barw* jest odpowiednia dla form fleksyjnych:

*podbarwiłem, podbarwiłeś, podbarwił, podbarwiła, podbarwiło, podbarwiliśmy, podbarwiliście, podbarwili, podbarwię, podbarwisz* itd.

Mając na uwadze względy praktyczne można utworzyć regułę przy użyciu krótszego ciągu znaków: *podb – pod|b*. Taka reguła ma zastosowanie dla wszystkich wyrazów rozpoczynających się od ciągu *podb*. W słowniku *sjp.pl* można znaleźć następujące formy podstawowe takich wyrazów:

*podbaldach, podbalkonowy, podbechtąć, podbechtywać, podbeletryzowany, podbeskidzki, podbiał, podbiała, podbicie, podbić, podbiec, podbieg, podbiegać, podbiegły, podbiegnąć, podbiegnięty, podbiegunowy, podbielać, podbielały, podbielić, podbieracz, podbierać, podbierak, podbieżeć, podbijać, podbijak, podbijarka, podbitka, podbojowy, podbora, podbój, podbramkowy, podbródek, podbródkowy, podbrudzać, podbrudzić, podbrygadier, podbrzeźniak, podbrzeże, podbrzusze, podbrzusznik, podbrzuszny, podbudowa, podbudować, podbudowywać, podbudówka, podbuntować, podbuntowywać, podburzać, podburzyć.*

Omówiony przykład obrazuje związek między długością ciągów liter wykorzystywanych w regułach a całkowitą liczbą reguł – skracanie długości tych ciągów pozwala na znaczną redukcję liczby reguł. Jednak krótsze ciągi znaków czasami występują również w wyrazach, które nie powinny być objęte regułami. Poniższa lista obejmuje wyrazy (formy podstawowe pobrane ze słownika sjp.pl), które rozpoczynają się od łańcucha *podz*:

*podzakres, podzamcze, podzamkowy, podzapytanie, podzastaw, podzbiornik, podzbiór, podzbiórka, podzelować, podzelowywać, podzespołowy, podzespół, podziabać, podziąć, podział, podzialać, podziałka, podziałowy, podziałówka, podziarno, podziczały, podziczyć, podzielać, podzielić, podzielnia, podzielnica, podzielniejszy, podzielnik, podzielność, podzielny, podziemie, podziemny, podziwiać, podziękować, podziękować, podziękowanie, podziobać, podziomek, podziobać, podziurawić, podziurkować, podziw, podziwiać, podziwić, podznaczenie, podzwaniać, podzwonić, podzwonne, podzwrotnikowy*

Łatwo zauważyć, że wymienione wyrazy rozpoczynają się od różnych przedrostków: *po* (np.: *podziobać, podzwonić*), *pod* (np.: *podzamkowy, podzapytanie*). Zatem zarówno reguła podziału: *podz – pod|z*, jak i reguła: *podz – po|dz* nie jest odpowiednia dla wszystkich wyrazów rozpoczynających się od łańcucha *podz*. Z tego wynika, że dla niektórych przedrostków trzeba stworzyć większą liczbę reguł opartych na dłuższych łańcuchach liter (reguł bardziej szczegółowych). Jednak w omawianym programie udostępniono jeszcze inną możliwość, która polega na definiowaniu zbioru reguł obejmujących wyjątki od danej reguły. Zatem można zdefiniować regułę *podz – pod|z*, która dotyczy wszystkich wyrazów rozpoczynających się od przedrostka *pod*. Jednocześnie można stworzyć listę reguł obejmujących wyrazy, dla których reguła *podz – pod|z* nie powinna być stosowana, np.: *podziob – po|dziob, podziób – po|dziób, podzwo – po|dzwo*.

Omówiony przykład jest specyficzny, ponieważ przedrostek *po* jest zawarty w przedrostku *pod*. Mechanizm tworzenia list reguł dotyczących wyjątków od danej reguły umożliwia uwzględnienie wyrazów, w których granica morfologiczna związana z obecnością przedrostka we współczesnej polszczyźnie nie jest jednoznaczna lub w ogóle nie istnieje. Niekiedy jedynie analiza historyczna może wyjaśnić strukturę morfologiczną niektórych wyrazów. W świadomości użytkowników języka niektóre złożenia zaczynają funkcjonować jako zwięzła całość. Dla takich wyrazów można zastosować mechanizm wyjątków, jednak trzeba podkreślić, że uwzględnianie w procesie dzielenia na sylaby istotności poszczególnych granic morfologicznych ma zawsze charakter umowny. Omawiane zagadnienia są szczególnie kłopotliwe dla wyrazów, których podział sugerowany przez morfologię i przez fonologię jest inny i nie można łatwo rozstrzygnąć, który z tych czynników jest bardziej istotny. W takich przypadkach strukturę reguł można oprzeć na wynikach testów dotyczących subiektywnego odczucia związanego z dzieleniem wyrazów na sylaby. Jako przykład kłopotliwych struktur morfologicznych można podać derywaty, dla których nie można wska-

zać jednej podstawy, ale wiążą się one z całą grupą leksemów współrzędnych (np. *zamknąć, odemknąć, domknąć*) (Grzegorzycykowa 1998: 365).

Z przedstawionych informacji wynika, że zestaw reguł podziału odnoszących się do płaszczyzny morfologicznej może przybierać różne kształty – reguły mogą być szczegółowe (tworzone dla konkretnych wyrazów) lub bardziej ogólne (oparte na krótszych łańcuchach liter). Poza tym liczba reguł zależy od liczby uwzględnionych granic morfologicznych. Trzeba również wspomnieć o trzech różnych podejściach do tworzenia zestawu reguł dotyczących struktury morfologicznej wyrazów. Pierwsze podejście zakłada, że zasady fonologiczne, skala sonorności oraz dodatkowe założenia są z góry zdefiniowane. Zatem zbiór reguł omówionych w rozdziale drugim jest określony. W takim przypadku dla płaszczyzny morfologicznej wystarczy zdefiniować tylko i wyłącznie te reguły, które nie są zgodne z regułami zdefiniowanymi dla płaszczyzny fonologicznej – musi istnieć możliwość przesłaniania reguł opartych na fonologii, jeżeli morfologia ma mieć wyższy priorytet w procesie dzielenia na sylaby. Natomiast podejście drugie nie przewiduje jakichkolwiek założeń dotyczących budowy reguł wynikających z zasad fonologicznych (dopuszczalne jest odniesienie do dowolnej skali sonorności oraz uwzględnienie dowolnych założeń). W takim przypadku należy utworzyć zestaw reguł uwzględniający wszystkie istotne granice morfologiczne (zbiór uwzględnionych granic morfologicznych jest określony). Podejście trzecie jest rozwinięciem podejścia drugiego. To rozwiązanie jest aktualnie rozwijane przez autora. Zakłada ono gromadzenie informacji o maksymalnej liczbie granic morfologicznych – nawet granic niezbyt wyraźnych lub dyskusyjnych. Poza tym rozwiązanie to przewiduje zapisywanie informacji o aktywności poszczególnych reguł. Zatem każda reguła może być w dowolnym momencie włączona lub wyłączona z procesu dzielenia na sylaby.

#### 4. Podsumowanie

W artykule omówiono program komputerowy przeznaczony do dzielenia wyrazów języka polskiego na sylaby. Najwięcej uwagi poświęcono strukturze reguł podziału wynikających z zasad fonologicznych oraz reguł opartych na morfologicznej budowie wyrazów. Rozwijany program umożliwia uwzględnienie tych dwóch czynników w dowolnym stopniu. Reguły obejmujące wewnątrzwyrazowe grupy spółgłoskowe są budowane na podstawie zasad fonologicznych, które w abstrakcyjny sposób odzwierciedlają zjawiska zachodzące w czasie artykułowania sylab. Przy tworzeniu listy reguł może być użyta dowolna skala sonorności, mogą być również przyjęte dowolne dodatkowe założenia lub nawet uwzględnione wyniki badań odczuć subiektywnych. Podobnie przy tworzeniu reguł dotyczących struktury morfologicznej wyrazów można uwzględnić określony zbiór przedrostków. Jeżeli morfologia ma znaczenie priorytetowe w procesie dzielenia na sylaby, to reguły oparte na strukturze morfologicznej będą przesłaniały reguły wynikające z zasad fonologicznych (reguły oparte na strukturze morfologicznej będą stosowane w pierwszej kolejności, a ewen-

tualne wstawienie znacznika granicy sylaby uniemożliwi zastosowanie w tym miejscu reguły wynikające z zasad fonologicznych). Planowane jest również stworzenie analogicznego mechanizmu dla przyrostków. Co prawda dla wyznaczania granic przyrostków potrzebna jest wiedza lingwistyczna, jednak badania odczuć subiektywnych wykazały, że istnieje pewna grupa często używanych formantów lub zakończeń (nie będących sufiksami), które mają istotny wpływ na decyzje o podziale na sylaby. Wyniki badań odczuć subiektywnych związane dzieleniem wyrazów na sylaby zostaną omówione w oddzielnej publikacji.

### Bibliografia

- DEMENKO Grażyna, WYPYCH Mikołaj, BARANOWSKA Emilia (2003): Implementation of Grapheme-to-Phoneme Rules and Extended SAMPA Alphabet in Polish Text-to-Speech Synthesis. – [w]: Maciej KARIPIŃSKI (red.): *Speech and Language Technology VII*. – Poznań: Polskie Towarzystwo Fonetyczne, 79–95.
- DOBROGOWSKA Katarzyna (1984): Śródgłosowe grupy spółgłosek w polskich tekstach popularnonaukowych. – *Polonica X*, 15–34.
- DOBROGOWSKA Katarzyna (1990): Word internal consonant clusters in Polish artistic prose. – *Studia Phonetica Posnaniensia II*, 43–67.
- DUNAJ Bogusław (1985): Grupy spółgłoskowe współczesnej polszczyzny mówionej (w języku mieszkańców Krakowa). – [w]: Adam HEINZ (red. serii): *Prace Językoznawcze DCCCIII*. – Kraków: nakł. Uniwersytetu Jagiellońskiego; Warszawa: PWN, 46–79.
- GRZEGORCZYKOWA Renata, LASKOWSKI Roman, WRÓBEL Henryk, red. (1998): *Gramatyka współczesnego języka polskiego. Morfologia*. – Warszawa: Wydawnictwo Naukowe PWN SA, 365.
- JASSEM Wiktor, ŁOBACZ Piotr (1974): Fonotaktyczna analiza mówionego tekstu polskiego. – *Biuletyn Polskiego Towarzystwa Językoznawczego XXXII*, 179–197.
- OSTASZEWSKA Danuta, TAMBOR Jolanta (2002): *Fonetyka i fonologia współczesnego języka polskiego*. – Warszawa: Wydawnictwo Naukowe PWN, 47–83.
- POLAŃSKI Edward, red. (2006): *Wielki słownik ortograficzny PWN z zasadami pisowni i interpunkcji*. – Warszawa: Wydawnictwo Naukowe PWN, 85–87.
- POLAŃSKI Kazimierz, red. (1999): *Encyklopedia językoznawstwa ogólnego*. – Wrocław: Zakład Narodowy im. Ossolińskich – Wydawnictwo, 644.
- SZPYRA-KOZŁOWSKA Jolanta (1998): The sonority scale and phonetic syllabification in Polish. – *Biuletyn Polskiego Towarzystwa Językoznawczego LIV*, 63–82.
- SZPYRA-KOZŁOWSKA Jolanta (2002): *Wprowadzenie do współczesnej fonologii*. – Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 150–152.
- STEFFEN-BATOGOWA Maria (1975): *Automatyzacja transkrypcji fonematycznej tekstów polskich*. – Warszawa: PWN, 78–90.
- ŚLEDZIŃSKI Daniel (2013): Podział korpusu tekstów na sylaby – analiza polskich grup spółgłoskowych. – *Kwartalnik Językoznawczy XV*, 48–100.
- TRASK Robert Lawrence (1996): *A dictionary of phonetics and phonology*. – New York: Routledge, 345.
- WIERZCHOWSKA Bożena (1971): *Wymowa polska*. – Warszawa: Państwowe Zakłady Wydawnictw Szkolnych, wyd. II, 214.