CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Przygotowanie i podział na sylaby korpusu tekstów dla dzieci.
Authors	Śledziński, Daniel;Szczerbinski, Marcin;Piotrowska, Joanna
Publication date	2013-01-01
Original Citation	Śledziński, D., Szczerbiński, M. and Piotrowska, J. (2013) 'Przygotowanie i podział na sylaby korpusu tekstów dla dzieci', in Scripta manent â res novae. Poznan : Wydawnictwo Naukowe UAM, pp. 439-450
Type of publication	Book chapter
Rights	© 2013, the Authors and Wydawnictwo Naukowe UAM. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-11-06 22:44:17
Item downloaded from	https://hdl.handle.net/10468/2957

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Daniel Śledziński
Instytut Językoznawstwa
Marcin Szczerbiński, Joanna Piotrowska
School of Applied Psychology, University College Cork

Przygotowanie i podział na sylaby korpusu tekstów dla dzieci

1. Wstęp

Artykuł dotyczy korpusu tekstów dla dzieci, który został utworzony na potrzeby projektu Graphogame-Fluent*, poświęconego ocenie skuteczności komputerowych gier edukacyjnych w terapii trudności w czytaniu (Szczerbiński et al. 2012). W artykule poruszono zagadnienia związane z przygotowaniem korpusu, przy czym najwięcej uwagi poświęcono podziałowi korpusu na sylaby (Śledziński 2010).

W publikacji przedstawiono kilka istotnych pojęć lingwistycznych – definicje: sylaby, sonorności i skali sonorności. Na podstawie dostępnych definicji nie da się przeprowadzić jednoznacznego podziału na sylaby. Problem ten w szczególności dotyczy języka polskiego – ze względu na występujące licznie wieloelementowe grupy spółgłoskowe o strukturze niespotykanej w innych językach. Na potrzeby praktyczne można jednak zastosować określone procedury i rozwiązania umowne. W artykule zaproponowano dwuetapową procedurę sylabizacji. Pierwszy etap obejmuje wyznaczenie kategorii zbitek spółgłoskowych oraz ustalenie dla tych kategorii wstępnego podziału opartego na zasadzie sonorności (jeżeli jest to możliwe). Etap drugi związany jest z ustaleniem umownych szczegółowych reguł podziału dla konkretnych zbitek spółgłoskowych przy uwzględnieniu wskazówek, które również zostały omówione w artykule.

2. Przygotowanie korpusu

W ramach projektu Graphogame-Fluent przygotowano korpus złożony z tekstów lektur szkolnych dla klas 1-6 szkoły podstawowej. W skład korpu-

* A science-based tool for training fluency in literacy for teachers and learners in English, Polish and German. EU Comenius Multilateral Project. Nr projektu: 510127-LLP-1-2010-FI--Comenius-CMP.

su miały wejść teksty wybrane na podstawie zestawienia kilkunastu list lektur szkolnych[1]. Ostatecznie do korpusu włączono 91 opracowań. Jest tam klasyka literatury dziecięcej (np. *Plastusiowy pamiętnik* M. Kownackiej) i teksty współczesne (np. *Najwyższa góra świata* A. Onichimowskiej), literatura polska (58 pozycji – np. *Król Maciuś Pierwszy* J. Korczaka) i światowa (33 pozycje – np. *Hobbit, czyli Tam i z powrotem* J.R.R. Tolkiena), poezja (8 pozycji – np. *Katechizm polskiego dziecka* W. Bełzy) i proza (83 pozycje – np. *W pustyni i w puszczy* H. Sienkiewicza).

W czasie gromadzenia korpusu napotkano na kilka problemów technicznych. Rozwiązanie tych problemów wiązało się z konwersją i ujednoliceniem kodowania plików tekstowych oraz korektą błędnie rozpoznanych znaków przez oprogramowania OCR (ang. *Optical Character Recognition*, tj. optyczne rozpoznawanie znaków).

Na podstawie korpusu utworzono listę frekwencyjną wyrazów (zawiera ona ponad 160 000 różnych form fleksyjnych wyrazów). Cały korpus podzielono na sylaby – dzięki temu można było wygenerować listę frekwencyjną sylab (znalazło się na niej ponad 11 000 różnych sylab). Wybrane sylaby z tej listy wykorzystano później w komputerowym treningu tempa czytania Graphogame-Fluent, z zastrzeżeniem, że żadna część korpusu nosząca cechy utworu nie będzie publikowana ani udostępniana. Korpus został zgromadzony w celu wykonania analiz lingwistycznych związanych z częstotliwością występowania określonych jednostek w tekstach.

3. Podstawowe pojęcia lingwistyczne

3.1. Definicje fonetyczne sylaby

Definicje fonetyczne odnoszą się do zjawisk fizycznych zachodzących w czasie artykulacji sylab (Dziubalska-Kołaczyk 2002). W *Encyklopedii językoznawstwa ogólnego* (Michowska, Wasielczyk 1999) zamieszczono następujące informacje o sylabie fonetycznej: „Odcinek wypowiedzi stanowiący jedność ekspiracyjną, ruchową i akustyczną, posiadający jedno maksimum donośności, który potencjalnie może być fonetycznie samodzielną wypowiedzią. Brak dotąd w pełni zadowalającej fonetycznej definicji sylaby". Dalej zwrócono uwagę na to, że sylaba była definiowana jako odcinek wypowiedzi między dwoma minimami: siły ekspiracji, energii artykulacyjnej, rozwarcia narządów artykulacyjnych oraz donośności. Podejście fonetyczne prezentuje

[1] Zgodnie z obowiązującą podstawą programową nie istnieje lista lektur obowiązkowych dla szkoły podstawowej. Na pierwszym etapie edukacyjnym nie wymienia się żadnych propozycji lektur, natomiast na drugim (klasy 4–6) podaje się przykładową listę lektur do wyboru.

też Bożena Wierzchowska w książce *Wymowa polska* (Wierzchowska 1971). Oto jak definiuje ona sylabę: „Odcinek mowy, zawarty między momentami jednoczesnych zmian w: układzie narządów mowy, ciśnieniu powietrza w tchawicy, natężeniu przebiegu akustycznego i jego donośności, nosi nazwę sylaby. Na ośrodek sylaby przypada maksimum rozwarcia narządów mowy, z czym wiąże się najniższe ciśnienie powietrza w tchawicy, najwyższe natężenie i największa donośność dźwięku. Na krańcach sylaby stopień zbliżenia narządów mowy jest znacznie większy, wzrasta też ciśnienie subglotalne, maleje za to natężenie i donośność dźwięku".

3.2. Definicje fonologiczne sylaby

Definicje fonologiczne opisują strukturę sylaby przy użyciu abstrakcyjnych klas dźwięków (samogłosek oraz spółgłosek). Z informacji zamieszczonych w *Encyklopedii językoznawstwa ogólnego* wynika, że sylaba złożona jest z obligatoryjnego ośrodka oraz fakultatywnych marginaliów, przy czym ośrodkiem sylaby jest najczęściej samogłoska. Marginalia sylaby obejmują nagłosową grupę spółgłoskową (inaczej następ sylaby) oraz wygłosową grupę spółgłoskową (inaczej zestęp sylaby). Sylaby pozbawione grupy wygłosowej nazywane są sylabami otwartymi. Niekiedy przyjmuje się dwudzielną strukturę sylaby z wyróżnieniem następu i rymu, przy czym rym zawiera obligatoryjny ośrodek sylaby i fakultatywny zestęp.

3.3. Sonorność i skala sonorności

Przypisywanie sonorności dźwiękom mowy można określić jako próbę przeniesienia na płaszczyznę fonologiczną niektórych fizycznych (fonetycznych) właściwości dźwięków. Sonorność dźwięków w sposób abstrakcyjny przedstawia ich donośność oraz właściwy dla nich stopień rozwarcia narządów artykulacyjnych (Trask 1996). Również Jolanta Szpyra-Kozłowska pisze wprost, że sonorność dźwięków określa stopień rozwarcia narządów mowy w trakcie ich artykulacji, jak również ich głośność. Poza tym autorka przedstawiła skalę sonorności właściwą dla języka polskiego (tabela 1) – skala ta została użyta w badaniach omówionych w niniejszym artykule (Szpyra-Kozłowska 2002).

Przytoczona skala sonorności nie jest jedyna – na przestrzeni lat zaproponowano kilka wariantów tej skali. W publikacji *The Sonority Scale and Phonetic Syllabification in Polish* autorka przytoczyła kilka bardziej znanych propozycji (Szpyra-Kozłowska 1998).

Tabela 1. Skala sonorności użyta w badaniach

Klasa głosek	Skrót	Sonorność	Zapis ortograficzny głosek
Samogłoski	Sa	6	a, e, i, o, y, u
Półsamogłoski	Pó	5	ł, j
Spółgłoski płynne	Pł	4	l, r
Spółgłoski nosowe	N	3	m, n, ń
Spółgłoski szczelinowe	Sz	2	f, w, s, z, sz, ż, rz, ś, ź, h, ch
Spółgłoski zwarte: zwarto-wybuchowe oraz zwarto-szczelinowe	Zw	1	p, b, t, d, k, g, c, dz, cz, dż, ć, dź

Z przedstawionymi właściwościami strukturalnymi sylaby oraz z pojęciem sonorności związana jest zasada sonorności (ang. Sonority Sequencing Principle), która zakłada, że sonorność dźwięków wzrasta w kierunku ośrodka sylaby i maleje z dala od niego. Zatem zasada sonorności w sposób abstrakcyjny odnosi się do zjawisk fonetycznych związanych z artykułowaniem sylab – między innymi do zmieniającego się stopnia rozwarcia narządów artykulacyjnych oraz do zmieniającej się donośności dźwięków. Z kolei „zasada maksymalnego nagłosu [...] powoduje przydzielenie do nagłosu następnej sylaby jak największej liczby spółgłosek, o ile taki podział nie jest sprzeczny z zasadą sonorności" (Szpyra-Kozłowska 2002). W przytoczonym opracowaniu sugerowane jest użycie tych zasad fonologicznych dla wyznaczania granic między sylabami. Również w przypadku korpusu tekstów dla dzieci użyto zasady sonorności oraz zasady maksymalnego nagłosu dla wyznaczania granic sylab, jednak omówione w następnym rozdziale problemy uwidaczniają, że dla języka polskiego zasady te nie są wystarczające.

W niniejszym artykule przyjęto założenie, że można wyznaczyć granicę między sylabami, opierając się na zasadzie sonorności, jeżeli istnieje takie miejsce (między dwiema głoskami), od którego sonorność wzrasta w kierunku zarówno poprzedzającego, jak i następnego ośrodka sylaby. Zatem założenie to nie dopuszcza występowania w nagłosie oraz w wygłosie sylaby również grupy spółgłoskowej złożonej ze spółgłosek o równej sonorności. Poza tym zakłada się, że pojedyncza spółgłoska ulokowana między ośrodkami sylab zawsze należy do nagłosu kolejnej sylaby (takie przypadki nie są omawiane w niniejszym artykule).

4. Wskazówki dotyczące sylabizacji

Zgodnie z fonologiczną zasadą sonorności, sonorność dźwięków powinna rosnąć w miarę zbliżania się do ośrodka sylaby i maleć w miarę oddalania się od niego. Wnikliwa analiza struktury grup spółgłoskowych języka polskiego ujawnia jednak dwa zasadnicze problemy:

– wewnątrz grupy spółgłoskowej co najmniej trójelementowej może nastąpić wzrost oraz spadek sonorności, zatem nie ma możliwości wyznaczenia takiego miejsca między głoskami, żeby sonorność wzrastała w obu kierunkach. Istnieją też inne struktury grup spółgłoskowych, których profil sonorności nie daje takiej możliwości (szczegółowe informacje znajdują się w rozdziale piątym),

– istnieje możliwość wyznaczenia granicy zgodnej z zasadą sonorności, jednak struktura morfologiczna wyrazu lub intuicja sugerują inny podział.

Wymienione problemy są często spotykane przy podziale grup spółgłoskowych wyrazów języka polskiego. Potwierdza to, że nie można opierać się tylko na zasadach fonologicznych przy wyznaczaniu granic sylab. Zatem wyznaczanie tych granic musi mieć charakter umowny, przynajmniej w określonych przypadkach. Nasuwa się pytanie czy jednostki wyznaczone na podstawie ustaleń umownych w dalszym ciągu można nazwać sylabami.

Pomimo wymienionych problemów, fonologia może stanowić solidny fundament dla sylabizacji wykonywanej dla celów praktycznych. Sylabizacja tekstów dla dzieci objęła dwa etapy. Najpierw, stosując zasadę sonorności, ustalono wstępne reguły podziału dla kategorii grup spółgłoskowych utworzonych na podstawie szerokiej klasy dźwięków (tam, gdzie było to możliwe – szczegółowe wyniki analiz przedstawiano w rozdziale piątym). Następnie odwołano się do różnych czynników, które umożliwiły przeprowadzenie podziału bardziej szczegółowego – uściślono granice dla poszczególnych grup spółgłoskowych lub nawet dla grup spółgłoskowych umiejscowionych w konkretnych wyrazach.

Istnieje przynajmniej klika czynników, które mogą okazać się pomocne przy wyznaczaniu szczegółowych reguł podziału – omówiono je w kolejnych podrozdziałach. Istotność poszczególnych czynników zależy od celu, dla którego wykonywany jest podział na sylaby tekstu lub nagranego sygnału mowy. Przedstawione wskazówki należy traktować jako praktyczne propozycje stanowiące uzupełnienie dla fonologicznej zasady sonorności, odnoszącej się w abstrakcyjny i uproszczony sposób do zjawisk fonetycznych i nie uwzględniającej struktury morfologicznej wyrazów. Wszystkie przedstawione propozycje mogą podlegać dyskusji oraz przyszłym modyfikacjom.

4.1. Wskazówki morfologiczne

Sylaby nie pokrywają się z morfami (Strutyński 2002). Jednak morfologia ma istotne znaczenie dla sylabizacji, ponieważ granica między morfami jest często odczuwalna jako naturalna granica sylab (w szczególności granica między prefiksem i morfem leksykalnym). Jednak w wielu przypadkach granica wynikająca z zasady sonorności nie pokrywa się z granicą morfolo-

giczną odczuwalną jako granica sylaby (np. podłokietnik, wystrzał, oszklona). Wykonując sylabizację na potrzeby praktyczne, trzeba zatem wybrać jedną opcję. Oczywiście kluczowe znaczenie ma istotność zachowania w konkretnym zastosowaniu granic morfologicznych lub fonetycznych (wyznaczonych przez fonologiczną zasadę sonorności).

4.2. Wskazówki fonetyczno-akustyczne

Widoczne na spektrogramie zjawiska fonetyczno-akustyczne są odzwierciedleniem procesu artykulacyjnego. Cyfrowy sygnał mowy można parametryzować, na przykład wyznaczając przebieg natężenia dźwięku lub iloczas głosek. Wiedza z zakresu fonetyki akustycznej może okazać się przydatna przy ustalaniu umownych granic sylab. Trzeba zaznaczyć, że fonologiczna skala sonorności w sposób abstrakcyjny przedstawia donośność głosek[2], która ma ścisły związek z natężeniem dźwięku. W szczególnych zastosowaniach fonetyka akustyczna może mieć znaczenie priorytetowe przy wyznaczaniu umownych granic sylab. Zastosowania te mogą być związane z podziałem nagranego sygnału mowy dla określonego zastosowania technicznego, na przykład dla budowy syntezatora mowy opartego na sylabach (Kishore, Black 2003, Shrikanth 2005, Kopecek 1999, Tatham et al. 1999). W takiej sytuacji potrzebna jest fachowa wiedza dotycząca spektrograficznego zapisu poszczególnych głosek, przejść pomiędzy głoskami czy zjawiska koartykulacji.

4.3. Wskazówki fonetyczno-artykulacyjne

Wskazówki fonetyczno-artykulacyjne mają związek z fonetyczną definicją sylaby. W teorii granica między sylabami powinna pokrywać się z maksymalnym zwarciem narządów artykulacyjnych. Jednak wyznaczenie jednego punktu na osi czasu (związanej z sygnałem mowy) może być kłopotliwe. W przypadku niektórych grup spółgłoskowych występuje seria zwarć (np.: łapka, kotka). Poza tym natura niektórych spółgłosek sugeruje ulokowanie maksimum zwarcia raczej w środkowej fazie ich artykulacji, a nie przed nimi (np. w przypadku spółgłoski szczelinowej otoczonej głoskami o wyższej sonorności). Wymienione problemy sprawiają, że nie da się polegać jedynie na wskazówkach fonetyczno-artykulacyjnych przy wyznaczaniu umownego podziału na sylaby, jednak wiedza dotycząca artykulacji

[2] Termin *donośność* został zdefiniowany przez B. Wierzchowską w *Wymowie polskiej*. Donośność dotyczy odczuć audytywnych właściwych dla poszczególnych głosek, natomiast natężenie dźwięku jest mierzalną cechą fizyczną.

dźwięków mowy może być bardzo przydatna. Trzeba zaznaczyć, że fonologiczna sonorność odnosi się również do stopnia rozwarcia narządów mowy, zatem w sposób abstrakcyjny przedstawia zjawiska fonetyczno-artykulacyjne.

4.4. Intuicja i odczucia subiektywne

Intuicja oraz odczucia subiektywne mogą mieć istotne znaczenie przy wyznaczaniu umownego podziału na sylaby. W języku polskim istnieje wiele grup spółgłoskowych, dla których podział nie jest klarowny, natomiast omawiane wcześniej zasady i wskazówki nie sugerują jednoznacznego rozwiązania. Największe znaczenie ma intuicja osoby lub członków zespołu wykonującego umowny podział na sylaby, jednak można posiłkować się też odczuciami innych osób, szczególnie tych, które nie mają fachowej wiedzy lingwistycznej (mogącej sugerować konkretne rozwiązania).

5. Analiza korpusu tekstowego

W rozdziale przedstawiono wyniki analizy korpusu tekstów dla dzieci. Wszystkie zbitki spółgłoskowe, które wystąpiły w korpusie przypisano do kategorii opartych na szerokich klasach dźwięków. Dla każdej kategorii ustalono sonorność poszczególnych spółgłosek (zgodnie z tabelą 1) – wyznaczono w ten sposób profil sonorności tych kategorii. Otrzymany zestaw danych pozwala na określenie przydatności fonologicznej zasady sonorności przy wyznaczaniu granic sylab w wyrazach języka polskiego.

Informacje zapisane w kolumnie Reguła utworzono w następujący sposób:

– litery V oraz C oznaczają samogłoskę (ang. *Vowel*) oraz spółgłoskę (ang. *Consonant*),

– pionowa kreska oznacza granicę dwóch kolejnych sylab,

– podano reguły podziału dla kategorii grup spółgłoskowych opartych na szerokich klasach dźwięków, ale tylko tam, gdzie pozwalała na to zasada sonorności (przy uwzględnieniu założenia zapisanego w podrozdziale 3.3),

– niekiedy była możliwość ulokowania granicy zgodnej z zasadą sonorności w dwóch miejscach (przy dłuższych grupach spółgłoskowych). W takich przypadkach stosowano zasadę maksymalnego nagłosu. Jako przykład można podać trójelementową grupę spółgłoskową, w której sonorność spółgłoski środkowej jest mniejsza niż sonorność spółgłosek marginalnych (np.: barwny, pulchny).

W dalszej części rozdziału przedstawiono wyniki analizy korpusu – w trzech tabelach zestawiono informacje o kategoriach grup spółgłosko-

wych: dwuelementowych, trójelementowych oraz czteroelementowych. Skróty użyte w kolumnie Kategoria oraz wartości sonorności zapisane w kolumnie Sonorność zostały już zdefiniowane w tabeli 1. Zgodnie z informacjami podanymi na wstępie, zapisane w tabelach reguły podziału dla poszczególnych kategorii zbitek spółgłoskowych mają charakter wyjściowy – ich ustalenie stanowi pierwszy etap sylabizacji.

5.1. Grupy spółgłoskowe dwuelementowe

W przypadku grup spółgłoskowych dwuelementowych zawsze można wyznaczyć granicę zgodną z zasadą sonorności (przy uwzględnieniu zasady maksymalnego nagłosu). Są tylko trzy możliwości:

$son(C_1) < son(C_2)$ -> $V|C_1C_2V$ (np. wy<u>pł</u>ata)

$son(C_1) > son(C_2)$ -> $VC_1|C_2V$ (np. ko<u>mp</u>uter)

$son(C_1) = son(C_2)$ -> $VC_1|C_2V$ (np. ko<u>mn</u>ata)

Grupy spółgłoskowe dwuelementowe występują w wyrazach języka polskiego znacznie częściej niż inne (dłuższe) grupy spółgłoskowe. Istnieje wiele przypadków dwuelementowych grup spółgłoskowych, dla których granice wyznaczone na podstawie zasady sonorności nie pokrywają się z intuicyjnie odczuwalną granicą wynikającą ze struktury morfologicznej wyrazu (np. granicą ulokowaną między prefiksem i morfem leksykalnym). Dlatego dla zaproponowanych w tabeli 2 reguł wyjściowych niezbędne może być wykonanie korekty uwzględniającej konkretne grupy spółgłoskowe lub konkretne wyrazy.

Tabela 2. Wstępne reguły podziału dla dwuelementowych grup spółgłoskowych

Lp.	Kat.	Son.	Reguła	Lp.	Kat.	Son.	Reguła
1	Zw-Zw	1-1	VC|CV	14	Pł-N	4-3	VC|CV
2	Zw-Sz	1-2	V|CCV	15	Pł-Pó	4-5	V|CCV
3	Zw-Pł	1-4	V|CCV	16	N-Zw	3-1	VC|CV
4	Zw-N	1-3	V|CCV	17	N-Sz	3-2	VC|CV
5	Zw-Pó	1-5	V|CCV	18	N-Pł	3-4	V|CCV
6	Sz-Zw	2-1	VC|CV	19	N-N	3-3	VC|CV
7	Sz-Sz	2-2	VC|CV	20	N-Pó	3-5	V|CCV
8	Sz-Pł	2-4	V|CCV	21	Pó-Zw	5-1	VC|CV
9	Sz-N	2-3	V|CCV	22	Pó-Sz	5-2	VC|CV
10	Sz-Pó	2-5	V|CCV	23	Pó-Pł	5-4	VC|CV
11	Pł-Zw	4-1	VC|CV	24	Pó-N	5-3	VC|CV
12	Pł-Sz	4-2	VC|CV	25	Pó-Pó	5-5	VC|CV
13	Pł-Pł	4-4	VC|CV				

5.2. Grupy spółgłoskowe trójelementowe

Jeżeli przyjąć, że zgodnie z zasadą sonorności w nagłosie sylaby sonorność może tylko rosnąć, a w wygłosie może ona tylko maleć (nagłos i wygłos sylaby nie mogą też zawierać ciągu spółgłosek o równej sonorności), to w badanym korpusie występuje wiele trójelementowych grup spółgłoskowych, dla podziału których zasady sonorności zastosować się nie da. Dotyczy to następujących sytuacji:

$son(C_1) < son(C_2)$ i $son(C_2) > son(C_3)$ (np.: znalazłbym)
$son(C_1) = son(C_2)$ i $son(C_2) > son(C_3)$ (np.: rozśpiewana)
$son(C_1) < son(C_2)$ i $son(C_2) = son(C_3)$ (np.: odwzajemnić)
$son(C_1) = son(C_2)$ i $son(C_2) = son(C_3)$ (np.: rozwrzeszczane)

W takich przypadkach, pomimo braku możliwości oparcia się na zasadzie sonorności, struktura morfologiczna wyrazu często podpowiada właściwe rozwiązanie. Z wymienionych w tabeli 3 problematycznych struktur trójelementowych (dla których nie można zastosować zasady sonorności), najczęściej są spotykane te, w których spółgłoska środkowa ma większą sonorność od pozostałych spółgłosek. W przypadku niektórych trójelementowych zbitek spółgłoskowych trudno wskazać najlepsze rozwiązanie (np.: punktualnie, zwierzchność). W takich sytuacjach trzeba polegać na intuicji i odczuciach subiektywnych, uwzględniając przy tym cel, dla którego wykonywany jest podział na sylaby.

Tabela 3. Wstępne reguły podziału dla trójelementowych grup spółgłoskowych

Lp.	Kat.	Son.	Reguła	Lp.	Kat.	Son.	Reguła
26	N-Zw-Pł	3-1-4	VC|CCV	61	Pł-N-N	4-3-3	VCC|CV
27	Zw-Zw-N	1-1-3	VC|CCV	62	Pł-Zw-Pł	4-1-4	VC|CCV
28	Zw-Sz-Pł	1-2-4	V|CCCV	63	Pł-Pó-Zw	4-5-1	
29	Zw-Sz-Zw	1-2-1		64	N-Zw-Zw	3-1-1	VCC|CV
30	Zw-Pł-N	1-4-3		65	N-Zw-Sz	3-1-2	VC|CCV
31	Zw-Zw-Sz	1-1-2	VC|CCV	66	N-Zw-Pó	3-1-5	VC|CCV
32	Zw-Zw-Pł	1-1-4	VC|CCV	67	N-Sz-Zw	3-2-1	VCC|CV
33	Zw-Zw-Zw	1-1-1		68	N-Sz-Sz	3-2-2	VCC|CV
34	Zw-Zw-Pó	1-1-5	VC|CCV	69	N-Sz-Pł	3-2-4	VC|CCV
35	Sz-Zw-Zw	2-1-1	VCC|CV	70	N-Sz-N	3-2-3	VC|CCV
36	Sz-Zw-Sz	2-1-2	VC|CCV	71	N-Sz-Pó	3-2-5	VC|CCV
37	Sz-Zw-Pł	2-1-4	VC|CCV	72	N-Zw-N	3-1-3	VC|CCV
38	Sz-Zw-Pó	2-1-5	VC|CCV	73	Zw-Sz-Sz	1-2-2	
39	Sz-Sz-Zw	2-2-1		74	Zw-Sz-N	1-2-3	V|CCCV
40	Sz-Sz-Sz	2-2-2		75	Zw-Sz-Pó	1-2-5	V|CCCV
41	Sz-Sz-Pł	2-2-4	VC|CCV	76	Zw-Pł-Zw	1-4-1	
42	Sz-Sz-N	2-2-3	VC|CCV	77	Zw-Pł-Sz	1-4-2	

43	Sz-Sz-Pó	2-2-5	VC|CCV	78	Zw-N-Pł	1-3-4	V|CCCV
44	Sz-Pł-Zw	2-4-1		79	Zw-N-Pó	1-3-5	V|CCCV
45	Sz-Pł-N	2-4-3		80	Zw-Pó-Zw	1-5-1	
46	Sz-N-Pł	2-3-4	V|CCCV	81	Pó-Zw-Zw	5-1-1	VCC|CV
47	Sz-N-N	2-3-3		82	Pó-Zw-Pó	5-1-5	VC|CCV
48	Sz-N-Zw	2-3-1		83	Pó-Sz-Zw	5-2-1	VCC|CV
49	Sz-Zw-N	2-1-3	VC|CCV	84	Pó-Sz-Sz	5-2-2	VCC|CV
50	Sz-Pó-Zw	2-5-1		85	Pó-Sz-Pł	5-2-4	VC|CCV
51	Pł-Zw-Zw	4-1-1	VCC|CV	86	Pó-Sz-N	5-2-3	VC|CCV
52	Pł-Zw-Sz	4-1-2	VC|CCV	87	Pó-Sz-Pó	5-2-5	VC|CCV
53	Pł-Zw-N	4-1-3	VC|CCV	88	Pó-Pł-Sz	5-4-2	
54	Pł-Zw-Pó	4-1-5	VC|CCV	89	Pó-N-Pł	5-3-4	VC|CCV
55	Pł-Sz-Zw	4-2-1	VCC|CV	90	Pó-N-N	5-3-3	VCC|CV
56	Pł-Sz-Sz	4-2-2	VCC|CV	91	Pó-N-Pó	5-3-5	VC|CCV
57	Pł-Sz-Pł	4-2-4	VC|CCV	92	Pó-Zw-Sz	5-1-2	VC|CCV
58	Pł-Sz-N	4-2-3	VC|CCV	93	Pó-Zw-Pł	5-1-4	VC|CCV
59	Pł-Sz-Pó	4-2-5	VC|CCV	94	Pó-Zw-N	5-1-3	VC|CCV
60	Pł-N-Zw	4-3-1	VCC|CV				

5.3. Grupy spółgłoskowe czteroelementowe

Z tabeli 4 wynika, że dla prawie połowy kategorii czteroelementowych grup spółgłoskowych nie można wskazać granicy podziału zgodnej z zasadą sonorności. Jednak takie grupy spółgłoskowe w tekstach występują sporadycznie, dlatego z punktu widzenia zastosowań praktycznych nie powinno to stanowić istotnego problemu. Poza tym dla większości czteroelementowych grup spółgłoskowych można wskazać granice sylab, odwołując się do struktury morfologicznej wyrazów (np. kontrwywiad, bezkształtny). Zdarzają się jednak przypadki, dla których trudno wskazać odpowiednie rozwiązanie.

Tabela 4. Wstępne reguły podziału dla czteroelementowych grup spółgłoskowych

Lp.	Kat.	Son.	Reguła	Lp.	Kat.	Son.	Reguła
95	Sz-Zw-Sz-N	2-1-2-3	VC|CCCV	112	N-Zw-Sz-Zw	3-1-2-1	
96	Sz-Zw-Sz-Zw	2-1-2-1		113	N-Sz-Zw-Sz	3-2-1-2	VCC|CCV
97	Sz-Zw-Zw-N	2-1-1-3	VCC|CCV	114	N-Sz-Zw-Pł	3-2-1-4	VCC|CCV
98	Sz-Sz-Sz-Zw	2-2-2-1		115	N-Sz-Zw-N	3-2-1-3	VCC|CCV
99	Sz-Sz-Sz- Pó	2-2-2-5		116	N-Zw-Zw-Pó	3-1-1-5	VCC|CCV
100	Sz-Sz-Zw-Sz	2-2-1-2		117	N-Zw-Pł-Sz	3-1-4-2	
101	Sz-Sz-Zw-Pł	2-2-1-4		118	N-Zw-Pł-Pł	3-1-4-4	
102	Sz-Pł-Sz-N	2-4-2-3		119	Zw-Sz-Sz-Zw	1-2-2-1	

103	Sz-Zw-Pł-Sz	2-1-4-2		120	Zw-Sz-Zw-Sz	1-2-1-2	
104	Sz-Zw-Zw-Pł	2-1-1-4	VCC|CCV	121	Zw-Sz-Zw-Pł	1-2-1-4	
105	Pł-Sz-Zw-Zw	4-2-1-1	VCCC|CV	122	Zw-Zw-Sz-N	1-1-2-3	VC|CCCV
106	Pł-Sz-Zw-Sz	4-2-1-2	VCC|CCV	123	Pó-Sz-Sz-N	5-2-2-3	VCC|CCV
107	Pł-Sz-Zw-N	4-2-1-3	VCC|CCV	124	Pó-Sz-Sz-Zw	5-2-2-1	VCC|CCV
108	Pł-N-Zw-Zw	4-3-1-1	VCCC|CV	125	Pó-Sz-Zw-Sz	5-2-1-2	VCC|CCV
109	Pł-Zw-Sz-Zw	4-1-2-1		126	Pó-Sz-Zw-Pł	5-2-1-4	VCC|CCV
110	Pł-Zw-Pł-N	4-1-4-3		127	Pó-Zw-Pł-Sz	5-1-4-2	
111	Pł-Zw-Pó-Zw	4-1-5-1		128	Pó-Zw-Zw-Pł	5-1-1-4	VCC|CCV

6. Podsumowanie

W artykule omówiono zagadnienia związane z przygotowaniem oraz sylabizacją korpusu tekstów dla dzieci. Korpus został utworzony dla analiz statystycznych związanych z doborem właściwych jednostek dla gier edukacyjnych dla dzieci. W artykule najwięcej uwagi poświęcono podziałowi na sylaby wyrazów korpusu – przytoczono kilka istotnych definicji oraz wykonano analizę wszystkich grup spółgłoskowych pod kątem możliwości zastosowania fonologicznej zasady sonorności przy podziale na sylaby. Wyniki analizy korpusu potwierdzają, że nie da się wykonać sylabizacji opartej wyłącznie na zasadzie sonorności. Jednak dla celów praktycznych można przyjąć określone rozwiązania umowne.

W artykule zaproponowano dwuetapowy proces sylabizacji wykonywanej dla celów praktycznych. Pierwszy etap polega na przypisaniu wszystkich grup spółgłoskowych do kategorii grup spółgłoskowych utworzonych przy użyciu szerokich klas dźwięków. Na tym etapie tworzone są wstępne reguły podziału na podstawie zasady sonorności (tam, gdzie jest to możliwe). Drugi etap polega na ustalaniu reguł szczegółowych – dla konkretnych grup spółgłoskowych lub nawet dla grup umiejscowionych w konkretnych wyrazach. Zatem reguły wstępne mogą w niektórych przypadkach ulegać modyfikacjom – w zależności od uwzględnionych wskazówek: morfologicznych, fonetyczno-akustycznych lub fonetyczno-artykulacyjnych. W niektórych sytuacjach kluczowe znaczenie może mieć intuicja i odczucia subiektywne osoby (lub osób) wykonujących podział na sylaby. Zaproponowany w artykule zestaw wskazówek nie jest ostateczny i może on podlegać dyskusji. Wskazówki te stanowią praktyczne uzupełnienie dla fonologicznej skali sonorności i zasady sonorności, bowiem skala sonorności przypisuje dźwiękom mowy abstrakcyjne wartości liczbowe, zatem nie oddaje ona wszystkich artykulacyjnych i akustycznych właściwości głosek. Natomiast zasada sonorności nie uwzględnia struktury morfologicznej wyrazów, która może mieć istotne znaczenie dla sylabizacji.

BIBLIOGRAFIA

Dziubalska-Kołaczyk, K. 2002. *Beats-and-binding phonology*. Frankfurt am Main: Peter Lang GmbH. 39-51.

Kishore, S. i A. Black. 2003. *Unit size in unit selection speech synthesis*. Geneva: Eurospeech. 1317-1320.

Kopecek, I. 1999. *Speech recognition and syllable segments*. Berlin–Heidelberg: Springer. 1-203.

Michowska, E. i K. Wasielczyk. 1999. *Encyklopedia językoznawstwa ogólnego*. Wrocław: Zakład Narodowy im. Ossolińskich – Wydawnictwo. 575.

Shrikanth, N. 2005. *Text to speech synthesis*. New Jersey: Pearson Education. 1-105.

Strutyński, J. 2002. *Gramatyka polska*. Kraków: Wydawnictwo Tomasz Strutyński. 62.

Szczerbiński, M., Piotrowska, J., Richardson, U., Keurulainen, A., Wierzchoń, P., Śledziński, D., i K. Landerl. 2012. *Online remediation of reading fluency problems: technology, pedagogy, psychology*. Cork: 6th European Conference on Game-Based Learning [poster].

Szpyra-Kozłoska, J. 1998. „The sonority scale and phonetic syllabification in Polish". *Biuletyn Polskiego Towarzystwa Językoznawczego* 1998/54. 63-82.

Szpyra-Kozłoska, J. 2002. *Wprowadzenie do współczesnej fonologii*. Lublin: Wydawnictwo UMCS. 147-155.

Śledziński, D. 2010a. „Analiza struktury grup spółgłoskowych w nagłosie oraz w wygłosie wyrazów w języku polskim". *Kwartalnik Językoznawczy* 2010/3-4.

Śledziński, D. 2010b. „Fonemy, difony, trifony i sylaby – charakterystyka jednostek na podstawie korpusu tekstowego". *Kwartalnik Językoznawczy* 2010/3-4.

Tatham, M., Lewis E. i K. Morton. 1999. *Syllable reconstruction in concatenated waveform speech synthesis*. San Francisco: Proceedings of the International Congress of Phonetic Sciences. 2303-2306.

Trask, R. 1996. *A dictionary of phonetics and phonology*. New York: Routledge. 327, 345.

Wierzchowska, B. 1971. *Wymowa polska*. Warszawa: Państwowe Zakłady Wydawnictw Szkolnych. 102-197 oraz 213-216.