

Daniel Śledziński

Fonemy, difony, trifony i sylaby – charakterystyka jednostek na podstawie korpusu tekstowego

1. Cel badań i założenia

Fonemy, difony, trifony i sylaby są używane w konkretnych rozwiązaniach praktycznych. Wybór odpowiedniej jednostki dla danego rozwiązania może nie być łatwy, ponieważ używanie poszczególnych jednostek ma swoje wady i zalety. Często z określonej właściwości, która stanowi niewątpliwą zaletę, wynika inna właściwość danej jednostki, która uniemożliwia albo utrudnia jej implementację. Dobór właściwych jednostek jest istotny dla wielu zastosowań i aplikacji, między innymi dla systemów automatycznego rozpoznawania mowy (ASR – ang. *Automatic Speech Recognition*), dla systemów syntezy mowy (TTS – ang. *Text To Speech*), przy projektowaniu struktur zbiorów nagrań dźwiękowych, przy projektowaniu testów testów percepcyjnych. Jeżeli ograniczyć rozważania tylko do systemów automatycznego rozpoznawania mowy i systemów syntezy mowy, to opierając się na literaturze dotyczącej implementacji tych systemów w różnych językach, można stwierdzić, że na przestrzeni lat podejmowano próby użycia wszystkich rozważanych jednostek. Próby te dotyczyły zastosowania:

- fonemów w systemach ASR¹,
- difonów w systemach ASR²,
- trifonów w systemach ASR³,
- sylab w systemach ASR⁴,

¹ R. Halavati, B.Sh. Saeed, H.Z. Saman, *Recognition of human speech phonemes using a novel fuzzy approach*, w: *Applied Soft Computing*, vol. 7, Issue 3, 2007, s. 828–839.

² Cz. Basztura, P. Lisiak, P. Staroniewicz, *Automatic speech recognition based on diphones*, w: *Electrotechnical Conference, MELECON 98, 9th Mediterranean*, Tel-Aviv 1998.

³ Al-Diri, Bashir and Sharieh, Ahmad and Qutiashat, *A speech recognition model based on tri-phones for the Arabic language*, w: *Advances in modelling. Series B: Signal processing and pattern recognition*, Munib 2007, s. 49–64.

⁴ I. Kopecek, *Speech Recognition and Syllable Segments*. Publisher Springer, Berlin-Heidelberg 1999, s. 1–203; N. Takahashi, S. Nakagawa, *Syllable Recognition Using Syllable-Segment Statistics and Syllable-Based HMM*, w: *7th International Conference on Spoken Language Processing*, Denver 2002.

- fonemów w systemach TTS⁵,
- difonów w systemach TTS⁶,
- trifonów w systemach TTS⁷,
- sylab w systemach TTS⁸.

Przytoczone źródła to jedynie niewielka część literatury, która dotyczy wykorzystania omawianych jednostek w systemach syntezy mowy oraz w systemach rozpoznawania mowy. Ponadto współczesne badania dotyczą dużej liczby języków. Odkrywane są również nowe obszary badawcze oraz nowe możliwości wykorzystania omawianych jednostek.

Zaprezentowane w artykule badania zostały wykonane dla języka polskiego. Celem tych badań było porównanie poszczególnych jednostek pod kątem następujących kryteriów:

- liczby różnych jednostek,
- struktury jednostek,
- pokrycia tekstów określoną liczbą jednostek,
- liczby różnych jednostek występujących w próbach losowych wyrazów,
- możliwości określania wyrazów na podstawie jednostek,
- możliwości określania wyrazów na podstawie par jednostek,
- łączliwości jednostek z jednostkami tego samego typu.

Przedstawione wyniki mają zastosowanie nie tylko praktyczne. Są one również istotne w wymiarze lingwistycznym. Przeprowadzone analizy porównawcze ujawniają wiele interesujących właściwości poszczególnych jednostek – niektóre badania z pewnością warto kontynuować, uwzględniając więcej szczegółów lub biorąc pod uwagę inne aspekty.

Analizy zostały wykonane na korpusie tekstowym liczącym dwa miliony wyrazów. Korpus zawiera teksty pochodzące z różnych źródeł – przede wszystkim z artykułów prasowych oraz z artykułów internetowych, a także z książek. Badano tylko poszczególne jednostki (fonemy, difony, trifony i sylaby) należące do kolejnych wyrazów korpusu – nie odnoszono się do treści

⁵ I. Mikuni, K. Ohta, *Phoneme based text-to-speech synthesis system*, w: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86*, vol. 11, 1986, s. 2435–2438.

⁶ M.M. Rashid, M.A. Hussain, M.S. Rahman, *Diphone preparation for Bangla text to speech synthesis*, w: *Computers and Information Technology, ICCIT '09*, Dhaka 2009, s. 226–230; D. Munkhtuya, K.P. Kuldip, *Diphone-Based Concatenative Speech Synthesis System for Mongolian*, w: *Proceedings of the International MultiConference of Engineers and Computer Scientists 2008*, Hong Kong 2008.

⁷ C.F.J. Huang, E. Graf, *Triphone based unit selection for concatenative visual speech synthesis* w: *Acoustics, Speech, and Signal Processing, ICASSP-93*, Minneapolis 1993.

⁸ E. Lewis, M. Tatham, *Word and syllable concatenation in text-to-speech synthesis*, w: *Sixth European Conference on Speech Communications and Technology, ESCA 1999*, s. 9615–9618.

zawartych w tekstach oraz do znaczenia poszczególnych wyrazów, fraz i zdań. Przed przystąpieniem do badań wszystkie wyrazy korpusu zostały wymieszane (każdy wyraz został umieszczony na losowej pozycji w korpusie). Korpus obejmuje tylko wyrazy złożone z minimum trzech grafemów. Badania objęły tylko i wyłącznie wewnątrzwyrazowe właściwości jednostek.

2. Podstawowe pojęcia

W niniejszym rozdziale zdefiniowano pojęcia i przedstawiono informacje istotne dla badań, których wyniki opisano w dalszym ciągu artykułu.

2.1. Definicja fonemu

Fonem jest najbardziej elementarną z porównywanych jednostek. W literaturze fonem jest zdefiniowany jednoznacznie⁹, dlatego nie ma potrzeby wprowadzania dodatkowych definicji. W badaniach do zapisywania poszczególnych jednostek została użyta transkrypcja SAMPA, która w przypadku języka polskiego obejmuje 39 fonemów wymienionych w Tabeli 1.¹⁰

Tabela 1. Klasy dźwięków i transkrypcja fonematyczna SAMPA

Klasa dźwięków mowy	Transkrypcja (SAMPA)
Spółgłoski zwarto-wybuchowe	p, b, t, d, k, g, c, J
Spółgłoski zwarto-szczelinowe	t^s, d^z, t^S, d^Z, t^s', d^z'
Spółgłoski szczelinowe	f, v, s, z, S, Z, s', z', x
Spółgłoski nosowe	m, n, n', N
Spółgłoski płynne	l, r
Półsamogłoski	w, j, w~, j~
Samogłoski	a, e, o, u, i, y

W Tabeli 2. przedstawiono dane dotyczące liczby wystąpień poszczególnych fonemów w korpusie złożonym z dwóch milionów wyrazów. Dane dotyczą sumarycznej liczby wystąpień (również takich samych fonemów w jednym wyrazie z korpusu), a nie liczby wyrazów zawierających poszczególne fonemy. Z przedstawionych informacji wynika, że w korpusie najczęściej występowały trzy samogłoski: /e/, /a/, /o/. Najczęściej występującą spółgłoską zwarto-wybuchową jest /t/, natomiast wśród spółgłosek szczelinowych jest to spółgłoska /s/. Najczęściej występującą półsamogłoską jest /j/, natomiast najczęściej występującą spółgłoską płynną jest /r/.

⁹ D. Ostaszewska, J. Tambor, *Fonetyka i fonologia współczesnego języka polskiego*, Warszawa 2000, s. 101.

¹⁰ G. Demenko, M. Wypych, E. Baranowska, *Implementation of Grapheme-to-Phoneme Rules and Extended SAMPA Alphabet in Polish Text-to-Speech Synthesis*, w: *Speech and Language Technology*, vol. 7, Poznań 2003, s. 79–95

Tabela 2. Lista frekwencyjna fonemów (SAMPA)

L.p.	Fonem	L.wyst.	L.p.	Fonem	L.wyst.	L.p.	Fonem	L.wyst.
1	e	12959448	14	k	404930	27	x	164106
2	a	1198785	15	m	401150	28	t^s'	161481
3	o	1158029	16	n'	348034	29	t^S	159419
4	t	604605	17	l	314448	30	Z	134795
5	y	595680	18	d	295672	31	w~	115022
6	j	574765	19	S	232413	32	c	93909
7	r	527788	20	w	225774	33	d^z'	72731
8	n	511629	21	z	208988	34	N	38025
9	i	484973	22	f	202121	35	d^z	29005
10	u	463426	23	t^s	198289	36	J	16191
11	p	462717	24	b	186620	37	z'	15804
12	s	445027	25	g	178680	38	j~	10732
13	v	433275	26	s'	173076	39	d^Z	3191

2.2. Definicja difonu

Difon (ang. *diphone*) nie jest jednostką jednoznacznie zdefiniowaną, nie jest to również jednostka językowa¹¹. Przy wymawianiu kolejnych głosek narządy artykulacyjne płynnie zmieniają swoje położenie od pozycji właściwej dla pierwszej głoski do pozycji właściwej dla drugiej głoski. To przejście pomiędzy sąsiadującymi głoskami (ang. *transient*) jest w fonetyce akustycznej niezwykle istotne, dlatego difon często jest utożsamiany właśnie z tym przejściem. Przejście pomiędzy sąsiadującymi głoskami może być analizowane od części centralnej jednej głoski do części centralnej drugiej głoski – w ten sam sposób może być definiowany difon. Difon może być też rozumiany jako dwie sąsiadujące głoski – ten sposób definiowania jest odpowiedni dla omawianych w artykule badań, ponieważ taką definicję można przenieść na płaszczyznę fonologiczną. Można przyjąć, że difon to dwa fonemy, które reprezentują dwie sąsiadujące głoski. Zatem wyraz złożony z N fonemów będzie zawierał $N-1$ difonów, pod warunkiem, że $N \geq 2$. To twierdzenie opiera się na założeniu, że kolejne difony w wyrazie nachodzą na siebie (takie założenie przyjęto w badaniach opisanych w artykule). Zatem przykładowy wyraz złożony z pięciu fonemów $F_1F_2F_3F_4F_5$, stosując wyżej wymienione założenia, zawiera następujące difony: F_1F_2 , F_2F_3 , F_3F_4 , F_4F_5 . Takie podejście obejmuje tylko połączenia wewnątrzwyrazowe. W Tabeli 3. przedstawiono 30 najczęściej występujących difonów w analizowanym korpusie. W sumie w korpusie wystąpiło 899 różnych difonów.

¹¹ *Encyklopedia językoznawstwa ogólnego*, pod red. K. Polańskiego, Wrocław 1999, s. 268, 644.

Tabela 3. Lista 30 najczęściej występujących difonów

L.p.	Difon	L.wyst.	L.p.	Difon	L.wyst.	L.p.	Difon	L.wyst.
1	je	275764	11	on	101480	21	Se	77388
2	st	159643	12	ko	98849	22	li	76743
3	n'e	158281	23	ow~	92901	29	go	75172
4	po	152997	14	ej	92056	24	pS	74466
5	ov	150211	15	ja	86213	25	vj	73912
6	ra	124412	16	vy	81235	26	za	73779
7	na	123805	17	te	80812	27	le	72947
8	ro	120333	18	n'i	79908	28	em	72868
9	va	103721	19	ne	79580	29	ny	71784
10	ta	102152	20	ka	79284	30	al	71553

2.3. Definicja trifonu

W fonetyce jednym z fundamentalnych zagadnień jest wpływ na daną głoskę głosek bezpośrednio z nią sąsiadujących (zjawisko koartykulacji) – w taki sposób może być definiowany trifon (ang. *triphone*) – jako głoska znajdująca się w otoczeniu określonych głosek. Z punktu widzenia prezentowanych w artykule badań najistotniejsza jest możliwość jednoznacznego wyznaczenia trifonów w wyrazach. Jest to możliwe dzięki przyjęciu definicji analogicznej do definicji difonu (podrozdział 2.2). Zatem trifon to trzy fonemy reprezentujące trzy kolejno po sobie następujące głoski. W wyrazie trifony nachodzą na siebie, zatem każdy wyraz o długości N fonemów zawiera $N-2$ trifonów, przy założeniu że $N \geq 3$. Na przykład wyraz złożony z pięciu fonemów $F_1F_2F_3F_4F_5$, stosując powyższe założenia, zawiera następujące trifony: $F_1F_2F_3$, $F_2F_3F_4$, $F_3F_4F_5$. To podejście uwzględnia tylko trifony wewnątrzwyrazowe. W Tabeli 4. przedstawiono 30 najczęściej występujących trifonów w badanym korpusie tekstowym. W korpusie tekstowym wystąpiło w sumie 9779 różnych trifonów.

Tabela 4. Lista 30 najczęściej występujących trifonów

L.p.	Trifon	L.wyst.	L.p.	Trifon	L.wyst.	L.p.	Trifon	L.wyst.
1	ova	53957	11	nyx	27635	21	pro	24038
2	ego	50989	12	ont^s	27628	22	jes	23777
3	vje	50759	13	tur	27524	23	ost	23022
4	pSe	47129	14	jon	26904	24	ove	21979
5	sta	46751	15	aJo	26452	25	spo	21816
6	cje	35223	16	ktu	26386	26	yst	21785
7	est	33130	17	pra	26177	27	ovy	21597
8	jow~	32225	18	pSy	26081	28	ovj	21521
9	mje	31755	19	pje	24287	29	ale	20833
10	os't^s'	28483	20	ent	24044	30	an'e	20612

2.4. Definicja sylaby

Definicja difonu oraz definicja trifonu pozwalają na wyznaczenie tych jednostek w wyrazach w sposób jednoznaczny. W literaturze można spotkać różne definicje sylaby – zarówno na płaszczyźnie fonetycznej, jak i na płaszczyźnie fonologicznej. Z punktu widzenia tych badań istotna jest możliwość jednoznacznego podziału wyrazów na sylaby.

Definiowanie sylaby na płaszczyźnie fonetycznej polega na odwołaniu się do zjawisk o naturze fizycznej. Podanie takiej definicji jest zadaniem niezwykle trudnym. Istnieje kilka teorii odnoszących się do cech fonetycznych (fizycznych) sylab¹². Każde z tych podejść odzwierciedla w pewnym stopniu jakiś fragment rzeczywistości, jednak z pewnością żadne z nich nie jest w pełni wyczerpujące i uniwersalne. Teorie te związane są z następującymi zjawiskami fizycznymi:

- ekspiracją (ang. *chest-pulse theory*) – sylaba jest odcinkiem mowy ulokowanym pomiędzy dwiema przerwami ciągłości wydechu,
- stopniem rozwarcia narządów artykulacyjnych – sylaba definiowana jest jako pojedyncze rozwarcie (eksplozja) i zwarcie (implozja) traktu głosowego, na ośrodek sylaby przypada maksimum rozwarcia narządów artykulacyjnych,
- donośnością akustyczną dźwięków – (ang. *the prominence-theory*) – sylaba definiowana jest poprzez ulokowanie jej pomiędzy kolejnymi minimami donośności akustycznej, na ośrodki sylaby przypadają dźwięki o największej donośności,
- napięciem mięśniowym (energią artykulacyjną) – sylaba definiowana jest poprzez ulokowanie jej pomiędzy kolejnymi maksimami napięcia mięśniowego, na ośrodki sylaby przypadają minima napięcia mięśniowego.

W literaturze spotykane są dwa podejścia związane z fonologicznym opisem struktury sylaby¹³. Pierwsze podejście zakłada występowanie:

- obligatoryjnego ośrodka sylaby – w języku polskim w roli ośrodka sylaby może występować tylko samogłoska (z bardzo nielicznymi wyjątkami), natomiast w innych językach mogą to być również spółgłoski

¹² E. Michowska, K. Wasielec, *Encyklopedia językoznawstwa ogólnego*, Wrocław 1999, s. 575; B. Wierzchowska, *Wymowa polska*, Warszawa 1971, s. 102–197, 213–216; S. Chalker, E. Weiner, *The Oxford Dictionary of English Grammar*, New York 1994, s. 387; R.L. Trask, *A Dictionary of Phonetics and Phonology*, New York 1996, s. 327, 345; K. Dziubalska-Kołaczyk, *Beats-and-Binding Phonology*, Frankfurt am Main 2002, s. 44.

¹³ J. Strutyński, *Gramatyka polska*, Kraków 2002, s. 63; S. Chalker, E. Weiner, *op.cit.*, s. 387; *Kognitywne podstawy języka i językoznawstwa*, pod red. E. Tabakowskiej, Kraków 2001, s. 165–166.

sonorne lub nawet spółgłoski szczelinowe; dla języka polskiego zakłada się, że dany wyraz ma taką samą liczbę samogłosek oraz sylab;

- fakultatywnych marginaliów – są to dźwięki należące do sylaby i znajdujące się przed lub za ośrodkiem sylaby; dźwięki znajdujące się przed ośrodkiem sylaby nazywane są nagłosem sylaby, natomiast dźwięki znajdujące się za ośrodkiem sylaby nazywane są wygłosem sylaby; istnienie nagłosu oraz wygłosu nie jest warunkiem niezbędnym dla istnienia sylaby – sylaba może składać się z samego ośrodka; zarówno nagłos, jak i wygłos sylaby mogą składać się z jednej lub z większej liczby spółgłosek (z grupy spółgłoskowej).

Inny sposób fonologicznego opisu struktury sylaby zakłada jej podział na:

- nagłos – jest to fakultatywna spółgłoska lub grupa spółgłoskowa stojąca przed samogłoską,
- rym – obejmuje on obligatoryjny ośrodek oraz fakultatywny wygłos; wyrazy zakończone takim samym rymem rymują się.

Przytoczone definicje nie pozwalają na wykonanie jednoznacznego podziału na sylaby wszystkich wyrazów w języku polskim, ponieważ jest to język silnie spółgłoskowy (ze względu na występujące licznie wieloelementowe grupy spółgłoskowe). Jednak – przy uwzględnieniu pewnych umownych założeń – można użyć definicji fonologicznej dla segmentacji wyrazów na sylaby. Trzeba jednak stosować określone reguły podziału. Niestety tych reguł nie da się w pełni oprzeć na przesłankach lingwistycznych – trzeba stosować pewne rozwiązania umowne. Przyjęta definicja sylaby wygląda następująco:

$$(N) O (W),$$

gdzie O jest obligatoryjnym ośrodkiem sylaby (samogłoską), (N) jest fakultatywnym nagłosem, a (W) – fakultatywnym wygłosem sylaby. Nagłos oraz wygłos może obejmować jedną spółgłoskę lub grupę spółgłoskową. Poza tym przyjęto następujące założenia dotyczące wyodrębniania sylab:

- granice między wyrazami stanowią zawsze granice sylab,
- dla dwóch bezpośrednio następujących po sobie ośrodków sylab stosuje się zawsze następującą regułę podziału: O|O,
- przy połączeniu sylaby otwartej z sylabą o pojedynczym nagłosie stosuje się zawsze następującą regułę podziału: O|NO,
- jeżeli pomiędzy kolejnymi ośrodkami sylab znajduje się więcej niż jedna spółgłoska, to stosuje się odpowiednią regułę podziału; stworzona przez autora baza zbitek spółgłoskowych obejmuje propozycje reguł podziału dla ponad 2500 zbitek spółgłoskowych występujących w języku polskim; wiele reguł ma charakter umowny, jednak większość z nich

opiera się na konkretnych przesłankach (na przykład na przesłankach fonetyczno-akustycznych); przy opracowywaniu tego zestawu reguł autor kierował się praktycznymi potrzebami związanymi z segmentacją akustycznego sygnału mowy na sylaby.

Z przedstawionych informacji wynika, że podział wyrazów na sylaby w pewnym stopniu musi się opierać na rozwiązaniach umownych, a nie na definicjach lingwistycznych – dotyczy to wieloelementowych grup spółgłoskowych, których w języku polskim jest bardzo dużo. Powstaje zatem pytanie, czy wyznaczone w ten sposób jednostki można nazywać sylabami. Rozpatrując ten problem z lingwistycznego punktu widzenia – można mieć pewne wątpliwości. Jednak podstawowym celem tych badań była analiza jednostek pod kątem właściwości istotnych dla zastosowań praktycznych, dlatego autor zdecydował o nazywaniu wyznaczonych w ten sposób jednostek sylabami. Pewną alternatywą mogłoby być używanie terminu „jednostki sylabiczne”.

W Tabeli 5. przedstawiono trzydzieści najczęściej występujących sylab w korpusie złożonym z dwóch milionów wyrazów. Bez trudu można zauważyć, że są to w zdecydowanej większości sylaby typu CV (sylaby złożone ze spółgłoski i samogłoski). W korpusie wystąpiło 8669 różnych sylab.

Tabela 5. Lista 30 najczęściej występujących sylab

L.p.	Sylaba	L.wyst.	L.p.	Sylaba	L.wyst.	L.p.	Sylaba	L.wyst.
1	po	109073	11	s'e	54547	21	mi	37432
2	n'e	106721	12	ko	52893	22	ty	347179
3	o	81481	13	le	49174	23	n'a	37112
4	va	70264	14	ka	46053	24	mo	37052
5	na	67975	15	li	45213	25	te	36400
6	go	64327	16	do	43887	26	a	36106
7	za	63958	17	ro	43101	27	ra	33377
8	je	58164	18	u	42741	28	no	31167
9	vy	56427	19	ne	41355	29	da	31105
10	ta	56011	20	n'i	40273	30	to	30103

2.5. Definicja wyrazu

Zdefiniowania wymaga również pojęcie wyrazu, ponieważ spotykane w literaturze definicje są niejednoznaczne lub wykluczają się wzajemnie. Wydaje się, że najbardziej odpowiednie dla tych badań jest pojęcie wyrazu ortograficznego¹⁴ – czyli ciągu znaków ograniczonego spacjami. Taka definicja umożliwia łatwe wyznaczenie wyrazów w tekstach. Jednak opisane w artykule badania opierają się na transkrypcji fonematycznej (SAMPA), dlatego przed wszelkimi analizami wyrazy ortograficzne musiały być konwertowane

¹⁴ *Encyklopedia językoznawstwa ogólnego*, s. 644.

na zapis fonematyczny. Zatem pod pojęciem wyrazu w tym artykule był rozumiany wyraz ortograficzny zapisany przy użyciu fonemów. Przyjęcie takiej definicji powoduje, że różne formy fleksyjne wyrazu są uznawane za różne wyrazy (różne wyrazy ortograficzne przekonwertowane na zapis fonematyczny).

W artykule często używano sformułowań: „liczba wyrazów w korpusie” albo „odsetek wyrazów w korpusie”. Sformułowanie to obejmują fragmenty korpusu z powtarzającymi się wyrazami (identyczne wyrazy w tekstach powtarzają się). Jeżeli rozważany był fragment korpusu bez powtórzeń, to używano określenia „różne wyrazy”.

3. Struktura jednostek

Przedstawione w tym rozdziale dane dotyczą struktury poszczególnych jednostek (difonów, trifonów oraz sylab). Struktury zostały określone przy użyciu szerokich klas dźwięków: samogłosek (skrót: V – ang. *vowel*), spółgłosek półotwartych (skrót: S – ang. *sonorant* – tutaj bez samogłosek, które zostały ujęte oddzielnie) oraz spółgłosek właściwych (skrót O – ang. *obstruent*). W badaniu analizowano liczbę wystąpień poszczególnych struktur jednostek (difonów, trifonów i sylab) w korpusie złożonym z dwóch milionów wyrazów.

3.1. Struktura difonów

W Tabeli 6. przedstawiono informacje na temat częstości występowania określonych struktur difonów. W drugiej kolumnie tabeli wymieniono wszystkie struktury difonów określone przy użyciu szerokich klas dźwięków (samogłosek, spółgłosek półotwartych oraz spółgłosek właściwych). Trzecia kolumna zawiera dane dotyczące liczby różnych difonów o określonych strukturach, natomiast kolumna czwarta dotyczy liczby wystąpień w korpusie difonów o określonej strukturze. Pozycje w tabeli zostały posortowane malejąco według danych z kolumny czwartej. Z informacji zawartych w Tabeli 6. wynika, że w tekstach w języku polskim najczęściej występuje połączenie spółgłoski właściwej ze samogłoską oraz połączenie samogłoski ze spółgłoską właściwą. Również połączenie spółgłoski półotwartej z samogłoską oraz połączenie samogłoski ze spółgłoską półotwartą znalazły się na początku listy. Połączenie dwóch spółgłosek jest spotykane często, jednak takich sekwencji jest mniej niż połączeń z udziałem samogłoski. Jeżeli zsumować wszystkie kombinacje połączeń dwóch spółgłosek, to ich liczba przekroczy dwa miliony. Najrzadziej w języku polskim występuje połączenie dwóch samogłosek – w korpusie złożonym z dwóch milionów wyrazów zaobserwowano mniej niż 30 000 takich przypadków.

Łączliwość fonemów to zagadnienie istotne – jest ono przedmiotem nie tylko badań fonologicznych, ale również badań fonetyczno-akustycznych. Analizując spektrograficzny zapis sygnału mowy, można zauważyć, że granice pomiędzy kolejnymi głoskami nie zawsze są wyraźne i możliwe do określenia w sposób jednoznaczny. Jest to związane z rodzajem połączenia – na przykład wyraźnej granicy pomiędzy półsamogłoską a samogłoską w praktyce nie sposób wskazać, natomiast w połączeniu samogłoski i spółgłoski zwarto-wybuchowej granicę można wyznaczyć łatwo – na podstawie zwarcia, które jest wyraźnie widoczne na spektrogramie. Wyznaczanie granic głosek w sygnale mowy to zagadnienie istotne również w rozwiązaniach technicznych (na przykład w syntezie mowy oraz w automatycznym rozpoznawaniu mowy), ponieważ rozwiązania te opierają się na nagrany sygnale mowy z naniesionymi granicami czasowymi segmentów.

Tabela 6. Struktura difonów

L.p.	Struktura difonu	Liczba różnych difonów	Liczba wystąpień w korpusie
1	O+V	121	2812441
2	V+O	138	2284505
3	S+V	42	2123337
4	V+S	60	1674913
5	O+O	199	877310
6	O+S	122	778584
7	S+O	145	440727
8	S+S	44	170517
9	V+V	28	28924

3.2. Struktura trifonów

Tabela 7. zawiera dane dotyczące występujących w języku polskim struktur trifonów. Najczęściej spotykana struktura złożona jest z samogłoski otoczonej dwiema spółgłoskami właściwymi. Taki ciąg fonemów mógłby stanowić sylabę zamkniętą z pojedynczym nagłosem oraz pojedynczym wygłosem lub sylabę typu CV ulokowaną przed kolejną sylabą z przynajmniej pojedynczym nagłosem. Spółgłoska właściwa otoczona dwiema samogłoskami występuje również często – może to być spółgłoska ulokowana w nagłosie sylaby bezpośrednio za ośrodkiem sylaby otwartej. Liczba struktur trifonów oraz samych trifonów jest bardzo duża. Trifon jest często wybierany jako jednostka dla konkretnych rozwiązań, głównie dla systemów automatycznego rozpoznawania mowy, ponieważ fonem znajdujący się pozycji środkowej trifonu jest zawsze otoczony przez te same fonemy, zatem czynnik wpływu głosek sąsiednich na głoskę centralną jest stały.

Tabela 7. Struktura trifonów

L.p.	Struktura trifonu	Liczba. różnych trifonów	Liczba wystąpień w korpusie	L.p.	Struktura trifonu	Liczba różnych trifonów	Liczba wystąpień w korpusie
1	O+V+O	1672	1132416	14	V+S+S	175	126031
2	V+O+V	659	1111609	15	O+O+S	216	105055
3	O+V+S	881	1029832	16	S+O+O	205	80983
4	S+V+O	691	984485	17	O+O+O	143	58561
5	V+S+V	223	785911	18	S+O+S	226	43094
6	O+S+V	510	762779	19	O+V+V	182	18051
7	O+O+V	629	616523	20	V+V+O	233	16128
8	S+V+S	304	553683	21	V+V+S	129	10967
9	V+S+O	564	436320	22	S+V+V	94	10611
10	V+O+S	534	399645	23	O+S+S	46	9728
11	V+O+O	686	391826	24	S+S+S	14	4118
12	S+O+V	534	286534	25	O+S+O	26	3213
13	S+S+V	176	162472	26	S+S+O	20	655

3.3. Struktura sylab

W Tabeli 8. przedstawiono struktury sylab, które wystąpiły w badanym korpusie. Podobnie jak zaprezentowane wyniki dotyczące difonów oraz trifonów, struktury sylab zostały opisane przy użyciu szerokich klas dźwięków (samogłosek, spółgłosek półotwartych oraz spółgłosek właściwych). Uwagę zwraca bardzo duża liczba dopuszczalnych struktur – w rzeczywistości jest ona jeszcze większa, ponieważ w tabeli uwzględniono tylko te struktury, które wystąpiły w korpusie więcej niż 1000 razy. W badanym materiale najwięcej było sylab otwartych z pojedynczym nagłosem (połączenie jednej spółgłoski właściwej i samogłoski lub jednej spółgłoski półotwartej i samogłoski). W sumie sylab typu CV (połączenie spółgłoski i samogłoski) zaobserwowano ponad dwa i pół miliona. Bardzo często spotykana struktura sylaby to połączenie dwóch spółgłosek oraz samogłoski (sylaba otwarta z podwójnym nagłosem). Na przykład sylaby zbudowane ze spółgłoski właściwej, spółgłoski półotwartej oraz samogłoski wystąpiły prawie 430 000 razy. W materiale badawczym często występowały sylaby zamknięte z pojedynczym nagłosem i pojedynczym wygłosem. Sylaby złożone z jednej spółgłoski właściwej, z samogłoski oraz z jednej spółgłoski półotwartej wystąpiły ponad 400 000 razy. Ogólna zaobserwowana tendencja polega na tym, że czym bardziej złożona struktura sylaby, tym rzadziej ona występuje.

Tabela 8. Struktura sylab

L.p.	Struktura sylaby	Liczba. różnych sylab	Liczba wystąpień w korpusie
1	O+V	121	1640583
2	S+V	42	953967
3	O+S+V	440	429603
4	O+V+S	772	411648
5	O+0+V	473	297435
6	O+V+O	764	283224
7	S+V+O	337	194022
8	S+V+S	263	188910
9	V	5	185303
10	O+S+V+S	774	97864
11	O+S+V+O	844	72757
12	O+O+V+O	485	63250
13	O+O+V+S	655	51797
14	O+O+S+V	243	50501
15	V+S	44	42499
16	V+O	50	31340
17	S+V+O+O	81	30091
18	S+S+V	55	24136
19	O+O+O+V	99	19861
20	O+O+S+V+S	157	16094
21	S+S+V+S	57	14468
22	O+V+S+O	252	12482
23	S+S+V+O	48	9486
24	S+V+S+O	113	7783
25	O+O+O+V+S	80	7604
26	O+O+S+V+O	137	7568
27	O+V+O+O	143	7394
28	O+S+V+S+O	112	6440
29	O+S+S+V	43	4758
30	O+S+V+O+O	132	3346
31	O+S+O+V	23	2418
32	O+O+V+S+O	81	2313
33	O+O+O+S+V+S	11	2157
34	O+V+S+O+O	31	2003
35	O+S+S+V+S	35	1895
36	O+O+O+V+O	36	1711
37	O+O+V+O+O	49	1602
38	O+V+S+S	38	1551
39	S+V+O+S	15	1331
40	O+O+O+S+V	15	1241
41	V+O+O	23	1080

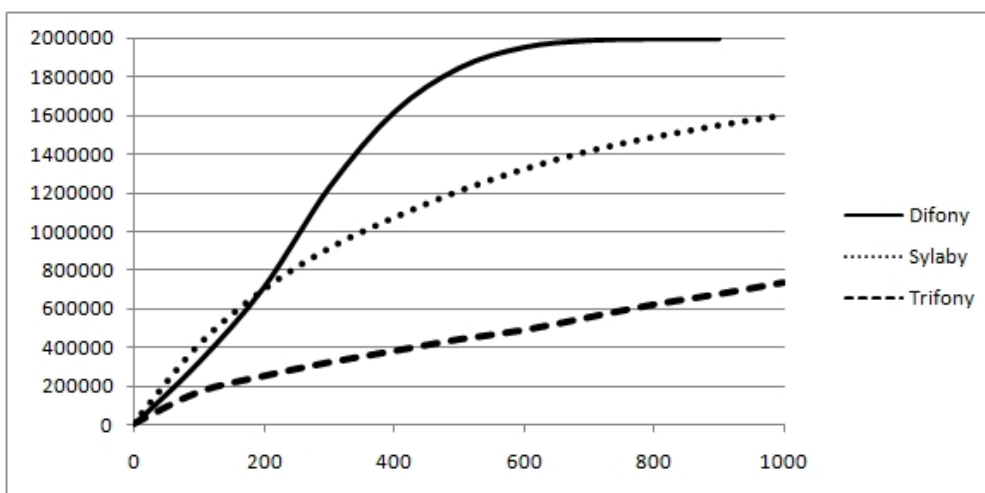
4. Pokrycie tekstu jednostkami

W tym rozdziale omówiono badanie, które dotyczy relacji pomiędzy liczbą różnych jednostek a liczbą wyrazów z korpusu, które można utworzyć z tych jednostek. Inaczej mówiąc, sprawdzano, jaki fragment korpusu można pokryć, dysponując określoną liczbą jednostek – difonów, trifonów oraz sylab. W badaniu nie uwzględniono fonemów – wynika to z oczywistego faktu, że stosunkowo niewielka liczba 39 fonemów wystarczy do utworzenia wszystkich wyrazów z korpusu. Pierwszy etap badania polegał na utworzeniu list frekwencyjnych jednostek (osobno dla difonów, trifonów oraz sylab). Dla każdej jednostki obliczono liczbę wystąpień w korpusie tekstowym oraz prawdopodobieństwo wystąpienia w losowym wyrazie. Wyniki posortowano (osobno: difony, trifony i sylaby). Dysponując takimi listami, można było przystąpić do zasadniczej fazy badania. Polegała ona na stopniowym pobieraniu coraz większej liczby jednostek z początku określonej listy frekwencyjnej. Dysponując zbiorem jednostek z początku listy frekwencyjnej, można było sprawdzić, ile wyrazów w korpusie tekstowym było pokrytych przez jednostki z tego zbioru (ile wyrazów nie zawierało ani jednej jednostki spoza tego zbioru jednostek). Wystąpienia identycznych wyrazów w korpusie były liczone osobno.

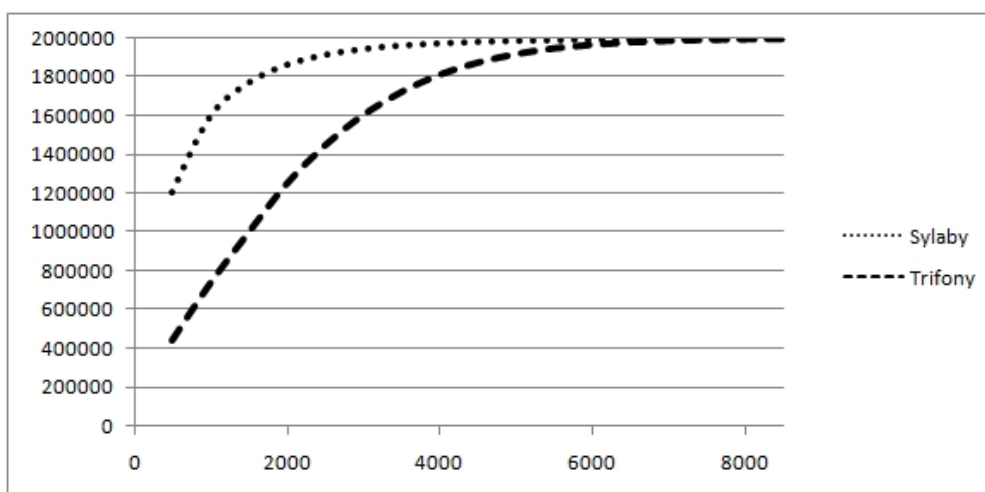
W pierwszej kolejności badanie wykonano dla liczby jednostek należącej do zakresu 0–1000 – sprawdzono pokrycie dla stu początkowych jednostek z poszczególnych list frekwencyjnych, a następnie dla każdej wielokrotności tej liczby (do tysiąca włącznie). Na Rysunku 1. zaprezentowano wynik badania. Z przedstawionych danych wynika, że stosunkowo niewielka liczba kilkuset najczęściej występujących jednostek (difonów, trifonów lub sylab) wystarczy dla pokrycia znacznych fragmentów korpusu tekstowego. Jednak wyniki uzyskane dla sylab oraz dla difonów są bardziej korzystne niż wyniki uzyskane dla trifonów. Na przykład, dysponując dwustoma najczęściej występującymi trifonami, można pokryć 248 928 wyrazów w korpusie, natomiast w przypadku sylab oraz difonów analogiczne wartości wynoszą odpowiednio 695 203 oraz 713 055. W miarę zwiększania liczby najczęściej występujących jednostek, zwiększa się różnica pokrycia pomiędzy sylabami i difonami. Mając do dyspozycji 500 najczęściej występujących trifonów, można pokryć 439 170 wyrazów w korpusie złożonym z dwóch milionów wyrazów. Odpowiednie wartości dla sylab i difonów to 1 205 619 oraz 1 848 540. Liczba 899 difonów pokrywa cały korpus tekstowy, ponieważ jest to liczba wszystkich (różnych) difonów. Natomiast dysponując liczbą tysiąca najczęściej występujących sylab i trifonów, można pokryć odpowiednio 1 598 749 oraz zaledwie 734 750 wyrazów w korpusie. Uzasadnione było zatem przeprowadzenie kolejnego badania – dla większej liczby jednostek pobieranych z list frekwencyjnych.

Kolejne omawiane badanie dotyczy tylko sylab oraz trifonów. Z list frekwencyjnych sylab i trifonów pobierano najpierw 500 początkowych elementów, a następnie wielokrotności tej liczby – aż do 8500 (czyli liczby zbliżonej do liczby różnych sylab). Wyniki przedstawiono na Rysunku 2. Z wykresu wynika, że 4000 najczęściej występujących sylab pokrywa niemal cały korpus (dokładnie 1 976 793 wyrazów). Dla trifonów podobny rezultat uzyskano dopiero przy około siedmiu tysiącach tych jednostek. Z przedstawionych danych

Rysunek 1. Pokrycie korpusu tekstowego jednostkami (w przedziale 0–1000)



Rysunek 2. Pokrycie korpusu tekstowego jednostkami (w przedziale 500–8500)



wynika, że najkorzystniejsze właściwości związane z pokryciem tekstów (mając na uwadze praktyczne zastosowania) mają difony, ponieważ stosunkowo niewielka liczba tych jednostek pokrywa znaczną część korpusu tekstowego. Wynik ten nie jest zaskakujący, ponieważ difonów jest wielokrotnie mniej niż sylab i trifonów. Istotne może być porównanie wyników uzyskanych dla sylab oraz trifonów – są one bardziej korzystne dla sylab. Wydaje się, że najważniejszy wynik tych analiz to liczby: 700 najczęściej występujących difonów, 4000 najczęściej występujących sylab oraz 7000 najczęściej występujących trifonów – użycie takich liczb jednostek pozwala na pokrycie niemal całego korpusu tekstowego.

5. Liczba jednostek w tekstach

Badania omówione w rozdziale czwartym dotyczą możliwości tworzenia wyrazów z określonego zbioru jednostek pobranych z początku listy frekwencyjnej (difonów, trifonów lub sylab). Ważne pytanie dotyczy tego, ile poszczególnych jednostek rzeczywiście występuje w tekstach. Dlatego przeprowadzono badanie, którego celem było sprawdzenie relacji pomiędzy liczbą wylosowanych wyrazów z korpusu a liczbą różnych fonemów, difonów, trifonów i sylab występujących w tej losowej próbie.

Na Rysunku 3. przedstawiono wykres zależności liczby różnych jednostek występujących w losowej próbie wyrazów od wielkości tej próby. Wyrazy były losowane z korpusu tekstowego zawierającego dwa miliony wyrazów. Wzięto pod uwagę liczbę losowanych wyrazów z zakresu od 100 do 1000, przy czym losowano najpierw 100 wyrazów, a następnie wielokrotność tej liczby (do 1000 włącznie). Po każdym losowaniu sprawdzano liczbę różnych jednostek (liczbę różnych fonemów, różnych difonów, różnych trifonów oraz różnych sylab) zawartych w próbie losowej. Z Rysunku 3. można odczytać, że 300 wylosowanych wyrazów zawierało mniej więcej równą liczbę difonów oraz sylab (około 400), jednak dalsze zwiększanie próby losowej skutkowało zwiększaniem przewagi liczby różnych sylab nad liczbą różnych difonów. Natomiast liczba różnych trifonów w każdej próbie losowej była znacznie większa niż liczba pozostałych jednostek. 1000 wylosowanych wyrazów zawierało prawie 2000 różnych trifonów.

Przedstawione na Rysunku 3. wyniki dotyczą stosunkowo niewielkiej liczby losowanych wyrazów (do 1000). Dane te są istotne i mają zastosowanie dla niewielkich tekstów. Jednak jeszcze ważniejsze są informacje o liczbie różnych jednostek w znacznie większych próbach losowych wyrazów. Na Rysunku 4. przedstawiono wyniki dotyczące liczby różnych difonów, różnych trifonów oraz różnych sylab w próbach losowych wyrazów, których liczebność mieści się w zakresie od 10 000 do 500 000. Krzywe dotyczące sylab oraz trifonów

zwracają uwagę podobnym kształtem, jednak krzywa dotycząca trifonów jest zamieszczona na wykresie wyżej. Dla badanego zakresu liczba trifonów jest za każdym razem (przy każdym losowaniu wyrazów) większa od liczby sylab o około 2000–2500. Największy wzrost liczby sylab oraz trifonów następuje przy próbach losowych wyrazów o liczebności mniejszej niż 100 000. Powyżej tej liczby tempo wzrostu liczby różnych jednostek stabilizuje się, jednak nie ustaje aż do liczby 500 000 losowych wyrazów. Wynika z tego, że dla uzyskania pełnej reprezentacji jednostek potrzebne są znacznie większe korpusy tekstowe (powyżej 500 000 wyrazów). Wydaje się, że dla kontynuowania tych badań trzeba odnieść się do wykazu zawierającego wszystkie formy fleksyjne wyrazów, a także trzeba dysponować znacznie większym korpusem tekstowym. Aktualne badanie pokazało, że 500 000 losowych wyrazów zawiera około 80% sylab oraz 90% trifonów, które wystąpiły w korpusie złożonym z dwóch milionów wyrazów.

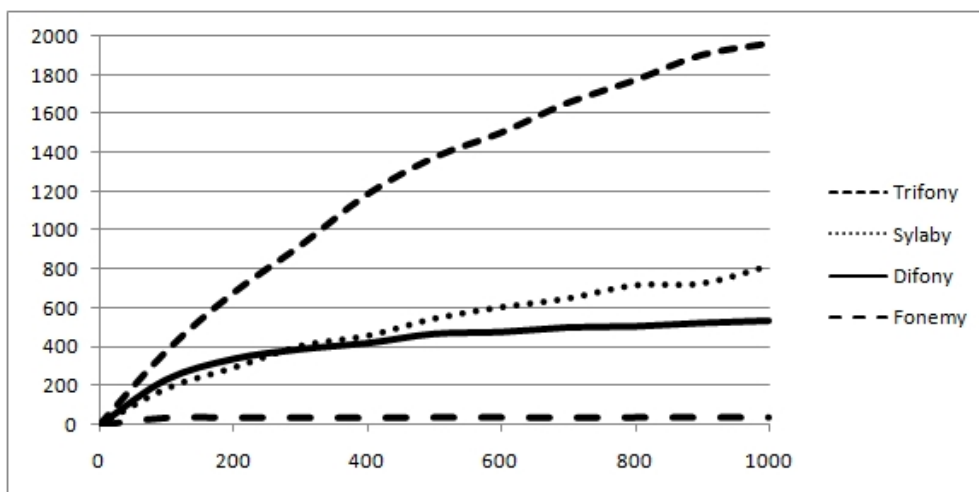
6. Określanie wyrazów na podstawie jednostek

W tym rozdziale przedstawiono wyniki badań, które dotyczą możliwości określania wyrazów na podstawie informacji o występowaniu jednostek. Obecność określonej jednostki wyznacza pewien skończony zbiór różnych wyrazów, które mogą zawierać tę konkretną jednostkę. Zagadnienie to jest istotne w systemach automatycznego rozpoznawania mowy, ponieważ ich działanie oparte jest na prawdopodobieństwie. Z punktu widzenia funkcji tych systemów najkorzystniejsza jest sytuacja, w której poszczególne jednostki są częściami jak najmniejszej liczby różnych wyrazów. Im mniejszy jest zbiór różnych wyrazów zawierających daną jednostkę, tym większe jest prawdopodobieństwo poszczególnych wyrazów z tego zbioru. To zagadnienie i uzyskane wyniki mogą mieć również inne zastosowanie – na przykład przy konstruowaniu testów percepcyjnych albo przy ustalaniu struktury zbiorów nagrań mowy.

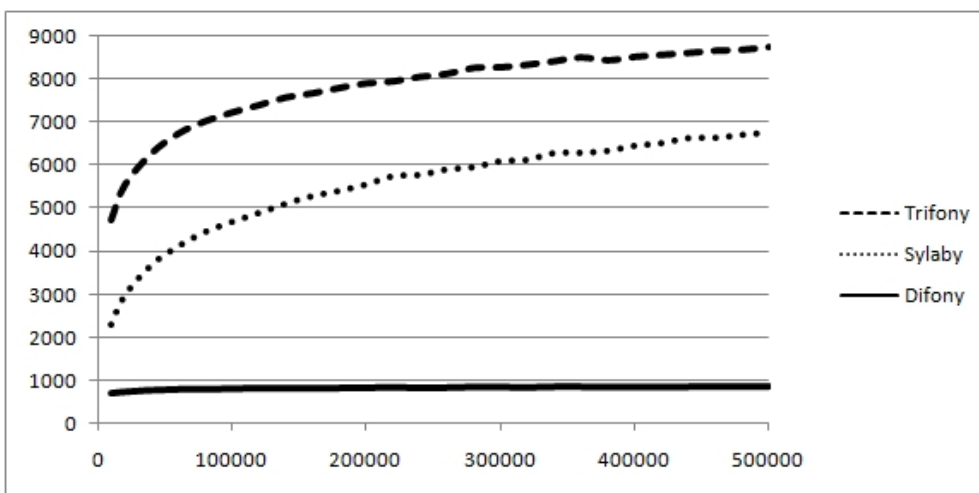
Na czterech histogramach przedstawiono wyniki badań kolejno dla: fonemów, difonów, trifonów oraz sylab. Z Rysunku 5. wynika, że obecność poszczególnych fonemów wyznacza bardzo dużo potencjalnych wyrazów, które mogą zawierać te fonemy. Na przykład 17,9% fonemów należy do liczby różnych wyrazów określonej przedziałem 40 001–50 000. Zatem możliwość określenia wyrazu na podstawie pojedynczego fonemu jest bardzo niewielka – wynika to z faktu istnienia niewielkiej liczby różnych fonemów.

Z badania przeprowadzonego dla difonów (Rysunek 6.) wynika, że około 30% difonów należy do najwyżej stu różnych wyrazów z korpusu, około 36% difonów należy do liczby różnych wyrazów określonej przedziałem 101–1000. Około 30% difonów jest częścią więcej niż 1000 różnych wyrazów w badanym

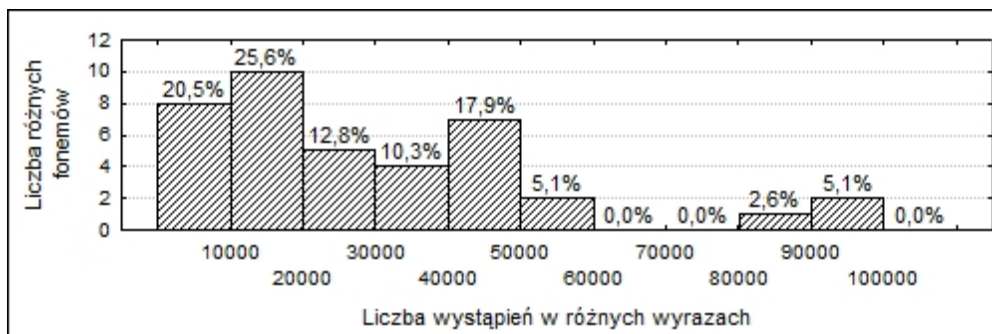
Rysunek 3. Zależność liczby różnych jednostek (oś Y) od liczby wylosowanych wyrazów (zakres 0–1000)



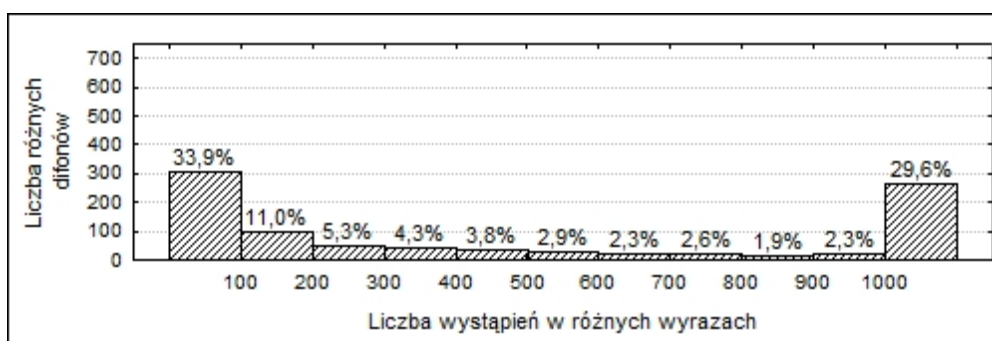
Rysunek 4. Zależność liczby różnych jednostek (oś Y) od liczby wylosowanych wyrazów (zakres 10 000–500 000)



Rysunek 5. Histogram obrazujący liczbę wystąpień fonemów w różnych wyrazach



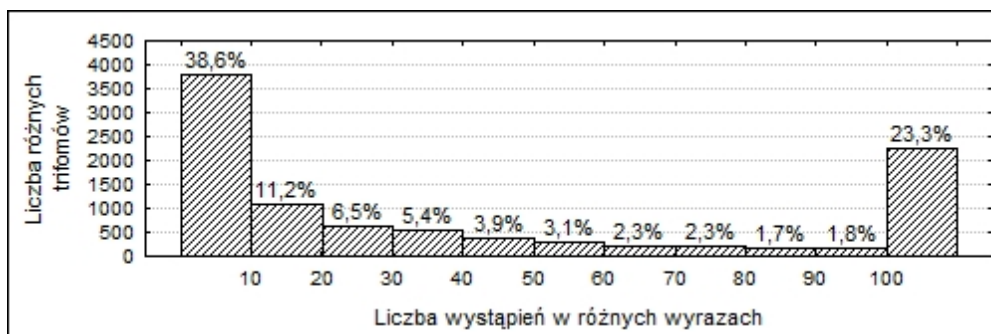
Rysunek 6. Histogram obrazujący liczbę wystąpień difonów w różnych wyrazach



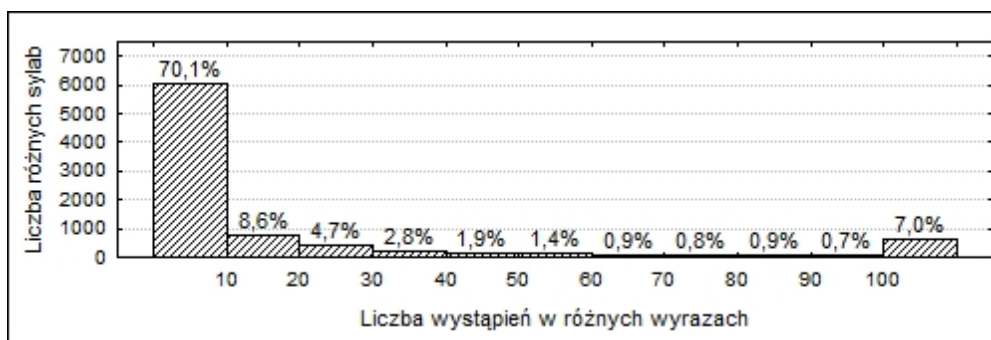
korpusie. Wyniki te są bardziej korzystne niż wyniki uzyskane dla fonemów, jednak możliwość określania wyrazów na podstawie pojedynczych difonów w dalszym ciągu nie jest duża, ponieważ większość difonów jest częścią więcej niż stu różnych wyrazów.

Wyniki uzyskane dla trifonów (Rysunek 7.) są znacznie bardziej korzystne niż wyniki uzyskane dla fonemów i difonów. Ponad 76% trifonów jest częścią mniej niż 100 różnych wyrazów, natomiast blisko połowa trifonów wystąpiła w mniej niż 20 różnych wyrazach. Jeszcze lepsze wyniki uzyskano dla sylab (Rysunek 8.) – około 70% sylab jest częścią mniej niż 10 różnych wyrazów w badanym korpusie złożonym z dwóch milionów wyrazów (analogiczny odsetek dla trifonów to 38%). Zatem z czterech porównywanych jednostek sylaby dają największe możliwości określania prawdopodobieństwa wyrazów.

Rysunek 7. Histogram obrazujący liczbę wystąpień trifonów w różnych wyrazach



Rysunek 8. Histogram obrazujący liczbę wystąpień sylab w różnych wyrazach



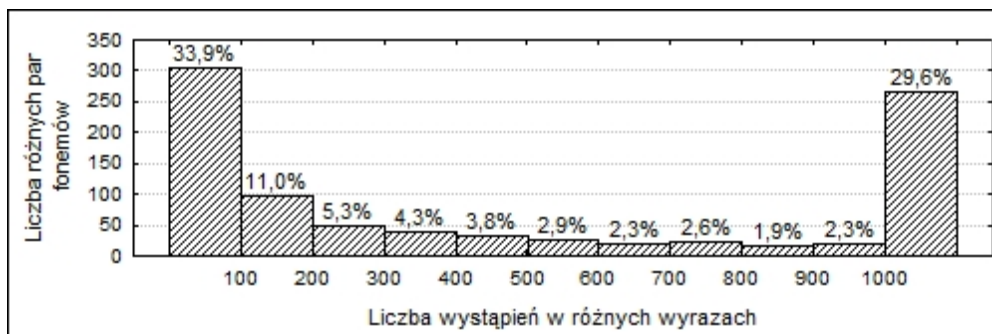
7. Określanie wyrazów na podstawie par jednostek

Opisane w rozdziale szóstym wyniki dotyczą możliwości określania wyrazów na podstawie poszczególnych jednostek (fonemów, difonów, trifonów oraz sylab). W niniejszym rozdziale prezentowane są wyniki dotyczące możliwości określania wyrazów na podstawie par jednostek (dwóch jednostek znajdujących się w bezpośrednim sąsiedztwie). Trzeba jednak zaznaczyć, że w przypadku difonów oraz trifonów przyjęte modele zakładają wzajemne nachodzenie na siebie jednostek, zatem pojęcie pary difonów lub pary trifonów oznacza parę difonów lub trifonów wzajemnie na siebie nachodzących.

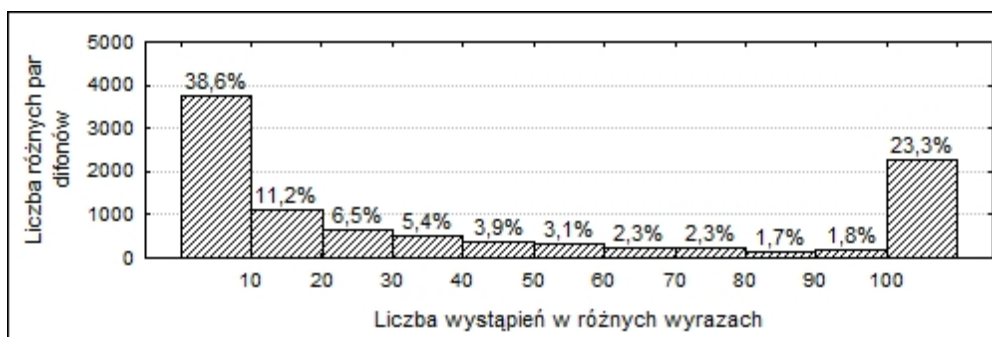
Na Rysunkach 9. i 10. przedstawiono histogramy z rezultatami badań dla fonemów oraz dla difonów. Wynika z nich, że większość par fonemów (ponad 70%) wystąpiło w więcej niż stu różnych wyrazach należących do korpusu tekstowego. W przypadku difonów analogiczny odsetek wyniósł 23,3%. Warto zauważyć, że ponad 38% par difonów jest częścią nie więcej niż 10 różnych wyrazów.

Interesujące wyniki uzyskano dla par trifonów oraz par sylab. Z informacji zamieszczonych na Rysunku 11. wynika, że 68% par trifonów wystąpiło w najwyżej 10 różnych wyrazach a zaledwie 3,1% par trifonów wystąpiło w więcej niż stu różnych wyrazach. Na Rysunku 12. przedstawiono histo-

Rysunek 9. Histogram obrazujący liczbę wystąpień par fonemów w różnych wyrazach



Rysunek 10. Histogram obrazujący liczbę wystąpień par difonów w różnych wyrazach

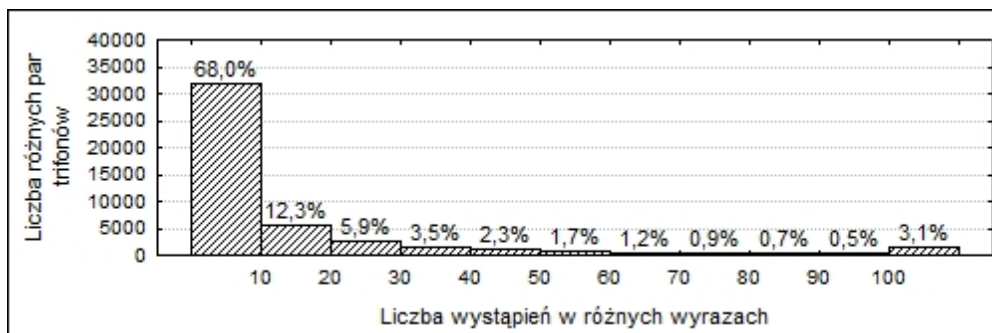


gram dotyczący liczby wystąpień par trifonów w różnych wyrazach, jednak wyszczególniony został zakres 1–10. Z histogramu wynika, że około 24% par trifonów wystąpiło tylko w jednym wyrazie, około 12% w dwóch różnych wyrazach, a około 8% w trzech różnych wyrazach z korpusu tekstowego.

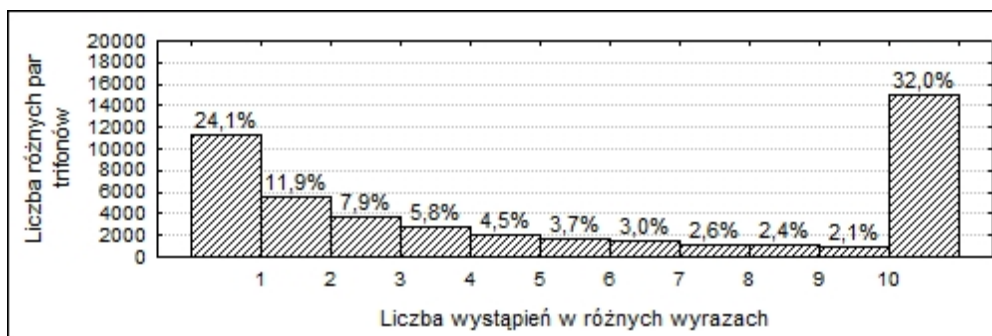
Najkorzystniejsze wyniki uzyskano dla sylab (dla par sylab). Z Rysunku 13. można odczytać, że aż 90% par sylab wystąpiło w najwyżej dziesięciu różnych wyrazach w korpusie. Kolejny histogram również dotyczy sylab, jednak wyszczególniono na nim zakres 1–10. Wynika z niego, że aż 53,4% par sylab wystąpiło tylko w jednym wyrazie, 14,6% par sylab wystąpiło w dwóch różnych wyrazach, a ponad 7% par sylab wystąpiło w trzech różnych wyrazach.

Podsumowując te badania, można stwierdzić, że sylaby są jednostkami najbardziej wyróżniającymi (wśród analizowanych jednostek) – w badanym korpusie 70% sylab wystąpiło najwyżej w dziesięciu różnych wyrazach, natomiast około 75% par sylab wystąpiło w najwyżej trzech różnych wyrazach (w korpusie złożonym z dwóch milionów wyrazów). W Tabeli 9. przedstawiono statystyki opisowe dotyczące możliwości określania wyrazów na podstawie jednostek oraz par jednostek. Uzyskane wartości średnie potwierdzają, że na podstawie sylab można wyznaczać wyrazy z największym prawdopodo-

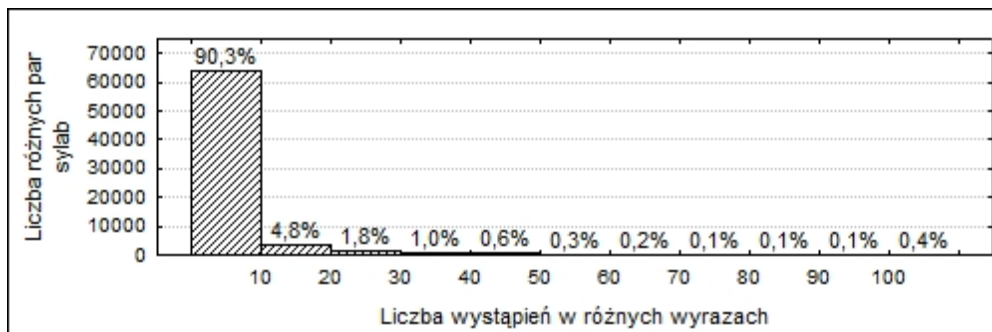
Rysunek 11. Histogram obrazujący liczbę wystąpień par trifonów w różnych wyrazach (zakres 0–100)



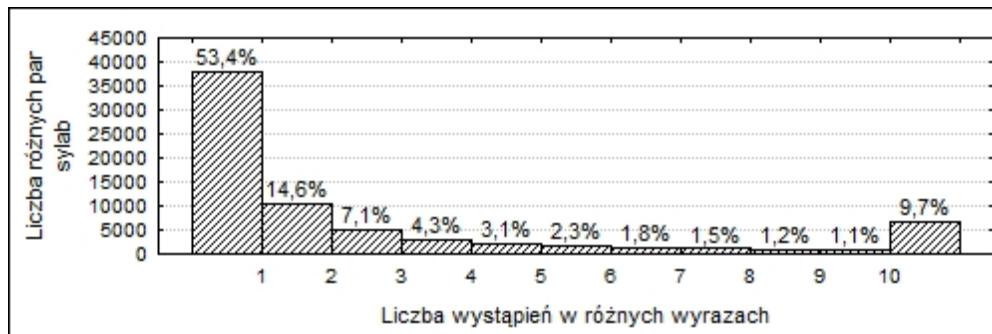
Rysunek 12. Histogram obrazujący liczbę wystąpień par trifonów w różnych wyrazach (zakres 0–10)



Rysunek 13. Histogram obrazujący liczbę wystąpień par sylab w różnych wyrazach (zakres 0–100)



Rysunek 14. Histogram obrazujący liczbę wystąpień par sylab w różnych wyrazach (zakres 0–10)



bieństwem. Statystyczna sylaba znalazła się średnio w prawie 60 różnych wyrazach w korpusie, natomiast statystyczna para sylab była częścią średnio 5,29 różnych wyrazów w korpusie.

Biorąc pod uwagę fakt, że w tym badaniu różne formy fleksyjne wyrazu były uznawane za różne wyrazy (różne wyrazy ortograficzne), rzeczywista możliwość wyznaczania wyrazów na podstawie sylab jest jeszcze większa. Aby uzyskać pełny obraz możliwości wyznaczania wyrazów na podstawie poszczególnych jednostek, należałoby wykonać badania uwzględniające dopuszczalne formy fleksyjne wyrazów, a także prawdopodobieństwo występowania poszczególnych jednostek. Celem przeprowadzonych badań było porównanie możliwości wyznaczania wyrazów w rzeczywistych tekstach na podstawie fonemów, difonów, trifonów oraz sylab, dlatego użyty materiał badawczy wydaje się być uzasadniony, a założony cel osiągnięty.

Tabela 10. Statystyki opisowe dotyczące liczby wystąpień jednostek i par jednostek w różnych wyrazach

Sekwencja	N ważnych	Średnia	Minimum	Maksimum	Odch. stand.
Pojedynczy fonem	39	28513,85	370	91686	23409,94
Pojedynczy difon	899	1264,56	1	22767	2402,12
Pojedynczy trifon	9779	102,30	1	10749	295,62
Pojedyncza sylaba	8669	59,77	1	13073	395,80
Para fonemów	899	1264,56	1	22767	2402,12
Para difonów	9795	102,13	1	10749	295,40
Para trifonów	46654	18,35	1	5393	64,21
Para sylab	70887	5,29	1	1520	20,74

8. Analiza łączliwości jednostek

W tym rozdziale przedstawiono wyniki badań dotyczących łączliwości poszczególnych jednostek z jednostkami tego samego typu w obrębie wyrazu. Analizowano liczbę różnych jednostek (różnych fonemów, różnych difonów, różnych trifonów i różnych sylab), które znalazły się w sąsiedztwie jednostek tego samego typu w korpusie złożonym z dwóch milionów wyrazów. Nie analizowano, jakie konkretnie jednostki stanowią to sąsiedztwo. Analizowano tylko i wyłącznie sąsiedztwo wewnątrzwyrazowe jednostek. Wyniki badania mogą mieć znaczenie w zastosowaniach praktycznych, na przykład w syntezie mowy, której działanie oparte jest na łączeniu (konkatenacji) określonych jednostek. Istotna może być także możliwość wyznaczania prawdopodobieństw jednostek potencjalnie sąsiadujących z daną jednostką. Badanie podzielono na dwie części – najpierw zbadano łączliwość lewostronną, a następnie łączliwość prawostronną fonemów, difonów, trifonów i sylab.

8.1. Analiza sąsiedztwa lewostronnego

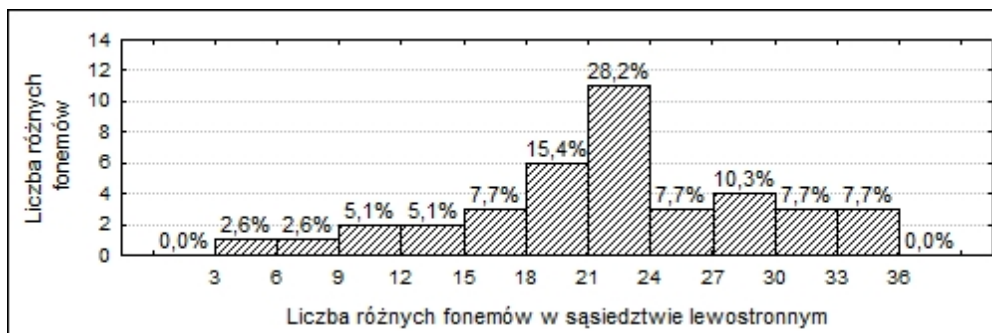
Na czterech histogramach przedstawiono wyniki badań dotyczących sąsiedztwa lewostronnego dla: fonemów, difonów, trifonów oraz sylab. Z informacji zamieszczonych na Rysunku 15. wynika, że dla większości fonemów liczba dopuszczalnych fonemów w sąsiedztwie lewostronnym mieści się w przedziale 18–30. Zaobserwowano 11 fonemów (28,2%), dla których liczba różnych fonemów w sąsiedztwie lewostronnym należy do przedziału 21–24.

Na Rysunku 16. zaprezentowano wyniki dotyczące sąsiedztwa lewostronnego difonów. Dla ponad połowy difonów zaobserwowano mniej niż 10 difonów w sąsiedztwie lewostronnym. Natomiast około 55% trifonów (Rysunek 17.) miało w swoim lewostronnym sąsiedztwie nie więcej niż 3 różne trifony. Około 66% sylab miało w swoim lewostronnym sąsiedztwie nie więcej niż 3 różne sylaby (Rysunek 18.).

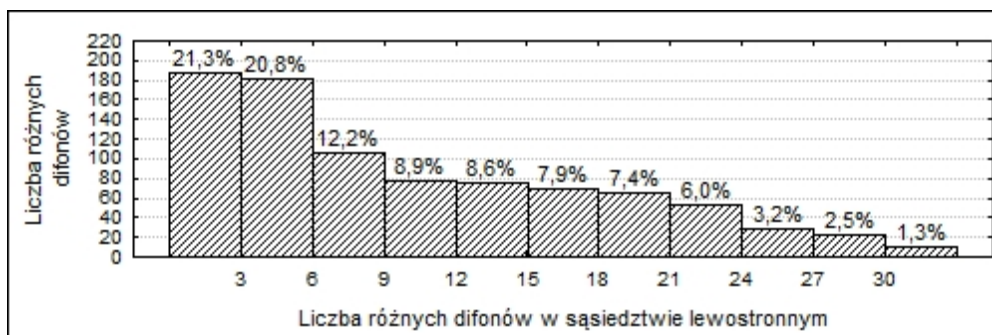
8.2. Analiza sąsiedztwa prawostronnego

Wyniki dotyczące sąsiedztwa prawostronnego fonemów są bardziej rozproszone niż wyniki uzyskane dla sąsiedztwa lewostronnego tych jednostek. Prawie 40% fonemów graniczy prawostronnie z liczbą fonemów należącą do przedziału 21–28. Prawie 16% fonemów graniczy prawostronnie z więcej niż 32 różnymi fonemami. Za to ponad 23% fonemów może graniczyć prawostronnie z najwyżej dwunastoma różnymi fonemami. 35% difonów może mieć w swoim sąsiedztwie prawostronnym najwyżej 4 różne difony. Analogiczny odsetek dla trifonów wyniósł 64,3% a dla sylab 55,4%.

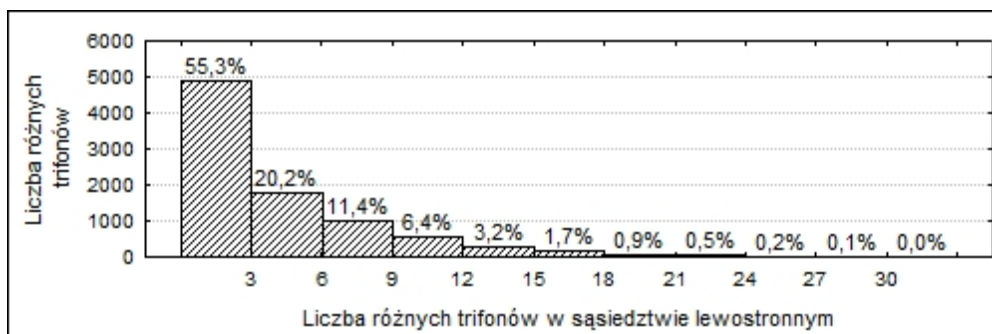
Rysunek 15. Sąsiedztwo lewostronne fonemów



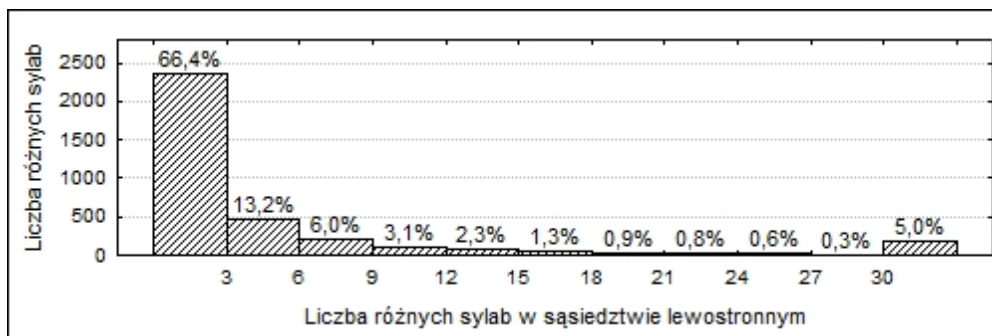
Rysunek 16. Sąsiedztwo lewostronne difonów



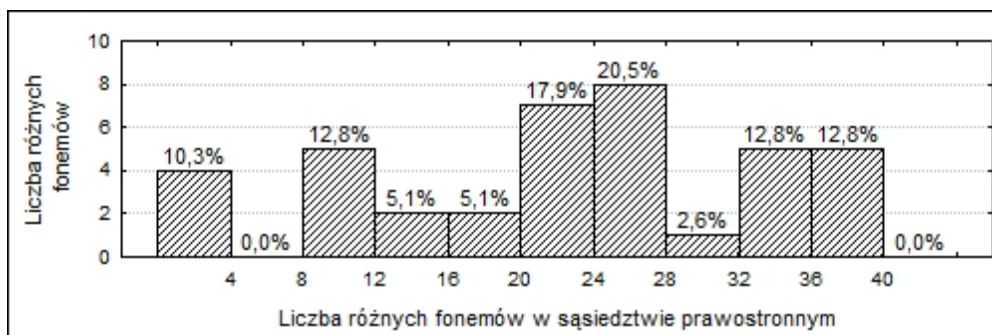
Rysunek 17. Sąsiedztwo lewostronne trifonów



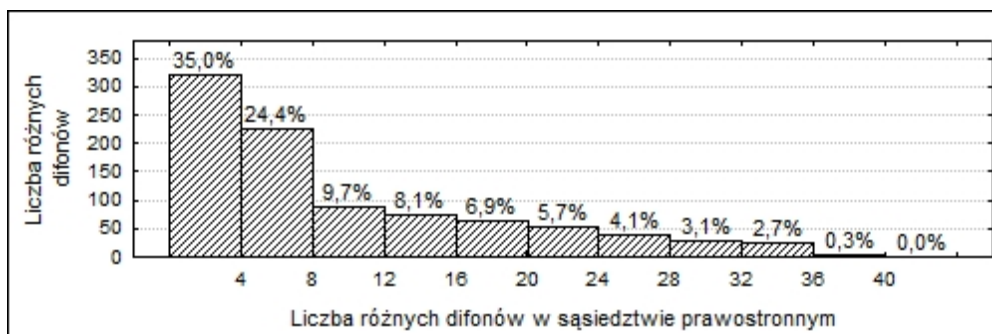
Rysunek 18. Sąsiedztwo lewostronne sylab



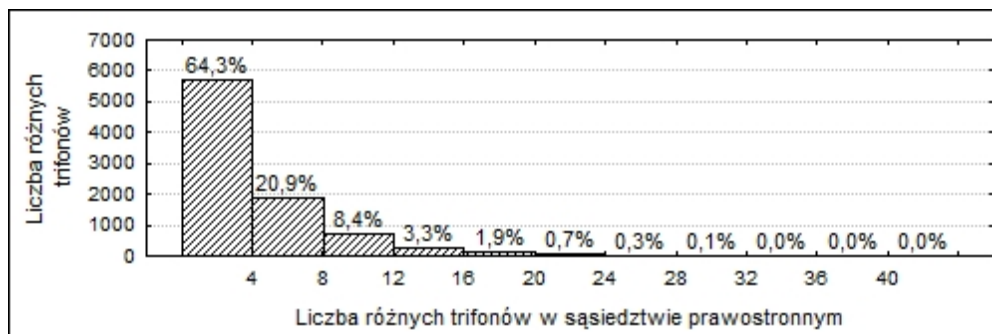
Rysunek 19. Sąsiedztwo prawostronne fonemów



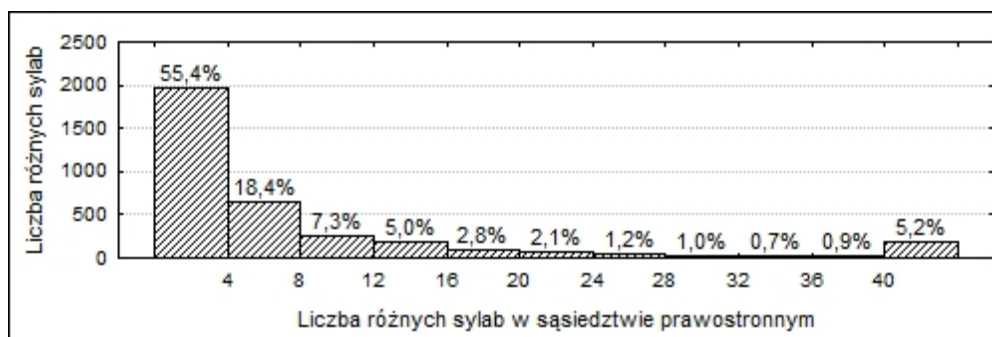
Rysunek 20. Sąsiedztwo prawostronne difonów



Rysunek 21. Sąsiedztwo prawostronne trifonów



Rysunek 22. Sąsiedztwo prawostronne sylab



Podsumowując badania związane z łączliwością fonemów, difonów, trifonów oraz sylab, można stwierdzić, że liczba jednostek, które mogą występować w sąsiedztwie danego typu jednostki, jest odwrotnie proporcjonalna do całkowitej liczby tego typu jednostek. Tendencję tą można zaobserwować, porównując średnie wartości liczby jednostek w sąsiedztwie poszczególnych jednostek – dane te zamieszczono w Tabeli 10. Fonemy, których jest najmniej (39), miały w swoim lewostronnym sąsiedztwie średnio 22,85 różne fonemy, a w sąsiedztwie prawostronnym bardzo zbliżoną liczbę – 22,74. W korpusie wystąpiło 899 różnych difonów i analogiczne wartości dla tych jednostek to 10,74 oraz 10,53. Zaobserwowana tendencja została zakłócona przez właściwości dystrybucyjne sylab – w badanym materiale wystąpiło niemal dziesięć razy więcej sylab niż difonów, jednak średnia liczba różnych sylab w sąsiedztwie lewostronnym sylab wyniosła 8,50, natomiast wartość dla sąsiedztwa prawostronnego jest większa niż wartość uzyskana dla difonów (11,92). Wy tłumaczeniem mogą być wartości w kolumnie Maksimum Tabeli 10. Dla sylab są one wielokrotnie większe niż dla pozostałych jednostek. Ponieważ wartości maksymalne dla sylab znacznie przekraczają wartości średnie, to można wnioskować, że istnieje stosunkowo niewielka liczba sylab, które mogą występować w sąsiedztwie dużej liczby innych sylab. Nietrudno odgadnąć, że

są to przede wszystkim sylaby typu CV. Natomiast liczba różnych trifonów była największa i zgodnie z zaobserwowaną tendencją średnia liczba różnych trifonów w sąsiedztwie lewostronnym trifonów wyniosła 4,64, natomiast w sąsiedztwie prawostronnym 4,60.

Tabela 10. Statystyki opisowe dotyczące sąsiedztwa jednostek

Jednostki	Sąsiedztwo	Średnia	Minimum	Maksimum	Odch. stand.
Fonemy	lewostronne	22,85	6	34	7,16
Difony	lewostronne	10,74	1	35	8,15
Trifony	lewostronne	4,64	1	30	4,51
Sylaby	lewostronne	8,50	1	479	27,14
Fonemy	prawostronne	22,74	2	38	10,86
Difony	prawostronne	10,53	1	38	9,16
Trifony	prawostronne	4,60	1	33	4,42
Sylaby	prawostronne	11,92	1	401	31,32

9. Podsumowanie badań

W artykule zaprezentowano wyniki badań, których celem było porównanie wybranych jednostek – fonemów, difonów, trifonów oraz sylab pod kątem określonych kryteriów. Difony oraz trifony nie są jednostkami językowymi. Fonem jest elementarną jednostką języka, natomiast sylaba jest jednostką związaną z procesem artykulacji. Pomimo tak znaczących różnic w definiowaniu lingwistycznych funkcji tych jednostek, łączy je to, że są one wykorzystywane dla celów praktycznych – w konkretnych rozwiązaniach i aplikacjach, między innymi w systemach syntezy mowy oraz w systemach automatycznego rozpoznawania mowy. Zatem uwzględnienie tych jednostek w badaniach było uzasadnione. Fonemy, difony i trifony bez trudu można wyznaczyć w fonematycznym zapisie wyrazów. Inaczej sytuacja wygląda w przypadku sylab – podział wyrazów na sylaby opiera się na rozwiązaniach umownych (w przypadku występowania grup spółgłoskowych pomiędzy ośrodkami sylab).

Omówione w artykule wyniki mogą być pomocne w ocenie przydatności poszczególnych jednostek dla konkretnych zastosowań. Poza tym zaprezentowane wyniki mogą stanowić punkt wyjścia dla bardziej szczegółowych badań lingwistycznych. Niektóre wyniki wydają się oczywiste i z góry do przewidzenia, na przykład liczba fonemów mniejsza niż liczba innych jednostek. Jednak wyniki te, umieszczone w kontekście wszystkich wyników uzyskanych dla wszystkich badanych jednostek, dają możliwość porównania i ewentualnie doboru najkorzystniejszej opcji dla konkretnego zastosowania.

Pierwszym porównywanym kryterium była liczba różnych jednostek w korpusie (różnych fonemów, różnych difonów, różnych trifonów oraz różnych sylab). Liczba fonemów była znana przed przystąpieniem do badań. Teoretyczna liczba połączeń difonowych to 1521, jednak badanie wykazało, że

w korpusie wystąpiło zaledwie 899 różnych difonów. Liczba różnych sylab w korpusie (8669) jest mniejsza od liczby różnych trifonów w tym samym materiale badawczym (9779).

Analiza struktury jednostek wykazała, że sylaby mają najbardziej różnorodną budowę – w korpusie występowały zarówno sylaby o nieskomplikowanej budowie (na przykład sylaby złożone tylko z samogłoski lub sylaby złożone ze spółgłoski oraz z samogłoski), jak i sylaby o budowie bardziej złożonej – na przykład sylaby zamknięte, które w swoim nagłosie i w wygłosie miały grupy spółgłoskowe. Struktura difonów oraz trifonów jest mniej różnorodna – wynika to z definicji (stałej długości) tych jednostek. Jednak liczba dopuszczalnych połączeń fonemów lub klas dźwięków wewnątrz difonów oraz trifonów jest bardzo duża.

Kolejne omawiane w artykule badanie miało na celu sprawdzenie, ile wyrazów w korpusie można pokryć, dysponując określoną liczbą jednostek pobranych z początku ich list frekwencyjnych. W badaniu nie uwzględniono fonemów, ponieważ 39 fonemów wystarczy do pokrycia wszystkich wyrazów. Również stosunkowo niewielka liczba 899 difonów pokrywa wszystkie wyrazy z korpusu. Istotne może być porównanie wyników uzyskanych dla trifonów i dla sylab. Badanie wykazało, że dysponując tysiącem najczęściej występujących sylab można pokryć około dwa razy więcej wyrazów niż w przypadku użycia takiej samej liczby trifonów. W miarę zwiększania liczby jednostek pobieranych z list frekwencyjnych ta proporcja się zmniejsza.

Zamierzeniem następnego badania było sprawdzenie, ile różnych jednostek znajdowało się w określonych próbach losowych wyrazów. Badanie wykazało, że w próbie losowej liczącej 1000 wyrazów jest najwięcej różnych trifonów (około 2000). W takiej próbie liczba różnych sylab (około 800) przewyższa liczbę różnych difonów (poniżej 600). Badanie powtórzono dla znacznie większego zakresu losowanych wyrazów. 500 000 wylosowanych wyrazów zawierało około 90% tych trifonów, które wystąpiły w całym korpusie złożonym z dwóch milionów wyrazów. Analogiczny odsetek dla sylab wyniósł około 80%.

Następne omówione badanie dotyczy możliwości określania wyrazów na podstawie jednostek. Badanie wykazało, że pojedyncze fonemy dają najmniejszą możliwość określania wyrazów, natomiast jednostką dającą największą możliwość określania wyrazów jest sylaba. Podobne rezultaty uzyskano dla badania, w którym analizowano możliwość określania wyrazów na podstawie par jednostek

Na końcu opisano wyniki badań dotyczących wewnątrzwyrazowej łączliwości poszczególnych jednostek z jednostkami tego samego typu. Osobno analizowano sąsiedztwo lewostronne oraz sąsiedztwo prawostronne poszczególnych jednostek. Brano pod uwagę tylko liczbę jednostek graniczących z da-

ną jednostką. Badanie wykazało, że fonemy mogą występować w sąsiedztwie największej liczby jednostek tego samego typu. Natomiast najmniejszą liczbę możliwych połączeń lewostronnych oraz prawostronnych wykazują trifony.

Omówione w artykule badania umożliwiają wstępną ocenę przydatności poszczególnych jednostek dla konkretnych zastosowań praktycznych. Badania pokazały również właściwości jednostek, które mogą być interesujące z punktu widzenia dalszych analiz lingwistycznych. Prawie każde omawiane w artykule badanie można kontynuować i rozwijać, na przykład poprzez analizę kolejnych szczegółów lub uwzględnienie innych aspektów.

**Phonemes, diphones, triphones and syllables
– characterising the units on the basis of corpus data**

SUMMARY

The purpose of this paper is to present the results of extensive investigations concerning selected phonological units: phonemes, diphones, triphones and syllables. The aim of the investigations was to compare these units, taking into consideration the following criteria:

- the number of units;
- the structure of units;
- the coverage of texts with a particular number of units;
- the number of units in randomly selected words from the corpus;
- the possibility of identifying words on the basis of particular units;
- the possibility of identifying words on the basis of particular pairs of units;
- the connectivity of units inside words.

The analysis is based on a text corpus containing two million words. The results achieved are useful both for specific practical solutions and for linguistic inquiries.

O Autorze

Daniel Śledziński - doktor nauk humanistycznych, Instytut Językoznawstwa, Uniwersytet im. Adama Mickiewicza w Poznaniu. Zainteresowania: przetwarzanie i analiza tekstów oraz sygnału mowy, percepcja mowy, fonetyka akustyczna, fonologia, języki programowania, tworzenie aplikacji, bazy danych, sztuczne sieci neuronowe, statystyka.
E-mail: danielssl@poczta.onet.pl