

Передмова до комп'ютерного фонетичного алфавіту SAMPA

Стрімкий розвиток мовних технологій, призначених для конверсії текст-на-озвучення, для автоматичного розпізнавання мовлення, для машинного перекладу, та ін. обумовив створення цілого ряду фонетичних алфавітів для комп'ютерного читання, які мають як локальне, так і міжнародне застосування.

До останніх можна віднести комп'ютерний (машинний) фонетичний алфавіт для відтворення складових мовленнєвого потоку SAMPA (Speech Assessment Methods Phonetic Alphabet), суть якого полягає в перекодуванні (mapping) символів Міжнародного фонетичного алфавіту IPA в 7-бітні символи ASCII в діапазоні 33..127.

Алфавіт SAMPA, а також його розширення X-SAMPA служать найвдалішою міжнародно узгодженою основою для стандартного комп'ютерного кодування фонетичних знаків, включаючи діакритики та позначки тонів, де кожен символ має своє унікальне призначення (uniquely parsable) і виводиться на екран з клавіатури без додаткових шрифтових розширень для фонетичних символів, що значно скорочує час, необхідний для запису транскрипції.

Ітак, всі IPA символи, які співпадають з малими літерами латиниці залишаються незмінними, всі інші знаки перекодовуються в ASCII символи в діапазоні 37..126.

Детальніше ознайомитися з історією створення комп'ютерного алфавіту SAMPA та його подальшого застосування для багатьох мов (наприклад, під егідою проекту COCOSDA) можна на сторінках:

<http://www.phon.ucl.ac.uk/home/sampa/home.htm>

<http://www.itl.atr.co.jp/cocosda/>

*Укладник комп'ютерного фонетичного алфавіту SAMPA для української мови **Тоня Білоус** запрошує до плідної співпраці всіх небайдужих до долі української мови та її місця у спільноті світових мов і прийме до уваги всі конструктивні зауваження та побажання. Таблиці подаються в порядку: українською мовою, англійською мовою.*

Пишіть мені за ел-адресами:

tonia.bilous@skynet.be

ukraspeech@skynet.be

*Author of the table "Ukrainian Sampa" (in Ukrainian and in English) - **Tonia Bilous***

You can kindly contact me by e-mail:

tonia.bilous@skynet.be

ukraspeech@skynet.be



Українська мова засобами фонетичного алфавіту *Sampa* (Speech Assessment Methods Phonetic Alphabet)

Основні приголосні фонери

У сучасній українській мові налічується 6 голосних фонем:

SAMPA символ	Транскрипція слова	IPA	Переклад англійською	Орфографія
A	rAk	rak	crayfish	рак
E	stEp	stɛp	steppe	степ
I	kIt	kɪt	whale	кіт
O	sOn	sɒn	dream	сон
u	ruX	ruX	movement	рух
i	k_j'it	kʲit	cat	кіт

Алофони голосних фонем

У ненаголошеній позиції, після напівприголосного /j/ або після пом'якшеного приголосного помітна невелика зміна в артикуляції голосних /A, u/. Вона стає трохи переднішою.

6	"zOn6	'zɒnɐ	zone	зона
U	tU"dI	tʊ'dɪ	there	туди

Ненаголошений /O/ перед наголошеним /u/ наближається у вимові до /u/ і стає /o/ , а ненаголошені /I/ та /E/ під впливом наступної наголошеної /E/ набувають характеристик /e/. Виникає гармонійна асиміляція.

o	to"mu	to'mu	because	тому
e	me"nE, me'nE	me'nɛ me'nɛ	will pass away, me	мине, мене

Різка зміна або послаблення артикуляції ненаголошених голосних звуків не є характерними для української мови.

Основні приголосні фонери

У сучасній українській мові налічується 32 приголосні фонери. Серед них 22 твердих приголосних. Для звуків /d, t, s, z, dz, ts, n, l/ характерна дорсально-зубна артикуляція:

SAMPA символ	Транскрипція слова	IPA	Переклад англійською	Орфографія
b	bIk	bɪk	bull	бик
p	pAr	pɑr	steam	пар

d	dAr	dar	gift	дар
t	tAm	tam	there	там
g	"grAtI	'gratɪ	bars	ґрати
k	kut	kut	angle	кут
h\	h\Ak	ɦak	hook	гак
x	xOr	xɔr	choire	хор
v\	v\Ir	vɪr	whirlpool	вир
f	fAx	fax	profession	фах
z	zAl	zal	hall	зал
s	sum	sum	sadness	сум
Z	ZAx	zax	horror	жах
S	Sum	ʃum	noise	шум
dz	"dzIg6	'dzɪgɐ	top	дзиґа
ts	tsE	tsɛ	this	це
dZ	dZm_jil'	dʒmʲilʲ	bumble-bee	джміль
tS	tSAs	tʃas	time	час
m	mEtS	mɛtʃ	sword	меч
n	nAs	nas	us	нас
r	rAb	rab	slave	раб
l	lAn	lan	field	лан

Також додаємо напівприголосний:

j	j6k	jek	how	як
---	-----	-----	-----	----

Сюди ж відносимо 9 м'яких (палаталізованих) приголосних, котрі виділені символом м'якості / ' / (ASCII 39) , який йде слідом за знаком приголосного. Для фонем / z', s', dz', ts'/ характерна дорсально-зубна артикуляція:

d'	"d'6d'ko	'dʲɛdʲkɔ	uncle	дядько
t'	t'Uk	tʲʊk	bale	тюк
z'	z'6t'	zʲɛtʲ	son-in-low	зять
s'	s'6d'	sʲɛdʲ	sit down	сядь
dz'	dz'Ob	dʒʲɔb	beak	дзьоб
ts'	ts'6	tsʲɐ	this	ця
n'	n'Ux	nʲʊx	scent	нюх
l'	l'Uk	lʲʊk	hatch	люк
r'	r'id	rʲid	generation	рід

Приголосні алофони

Є також 13 "напівпом'якшених" (напівпалаталізованих) приголосних, котрі розглядаються як алофони їх парних твердих фонем, і які помічені символом (тут) напівм'якості /_j /, який йде услід за знаком приголосного.

Так як в алфавіті IPA відсутні окремі символи, що могли б служити для розрізнення ступенів м'якості приголосних, а наявний лише загальний символ палаталізації [^j], то ним помічаються як м'які, так і напівпом'якшені приголосні. Це фонематично не несе двозначності, так як в українській мові немає пар приголосних, протиставлених за ознакою ступеню м'якості. (див. Примітки)

m_j	m_jid'	m^jid^j	copper	мідь
b_j	b_jik	b^jik	side	бік
p_j	p_jid	p^jid	under	під
g_j	g_jips	g^jips	gypsum	гіпс
k_j	k_jin'	k^jin^j	horse	кінь
x_j	x_jid	x^jid	motion	хід
h_j	h_jimn	h^jimn	anthem	гімн
v_j	v_jin	v^jin	he	він
f_j'	f_jin	f^jin	Finn	фінн
Z_j	"Z_jink6	'z^jinkə	woman	жінка
S_j	S_jis't'	ʃ^is^t^j	six	шість
dZ_j	bdZ_jil	bdʒ^jil	bees	бджіл
tS_j	tS_jip	tʃ^jip	plug	чіп

Інші алофони

Алофон / I_^ / називають нескладотворчим **голосним** варіантом фонем /i/ або ж нескладотворчим **приголосним** алофоном фонем /j/:

I_^	h_AI_^	hɑI_	grove	гай
-----	---------	------	-------	-----

Губний алофон /w/ губно-зубного сонорного приголосного / v_ / реалізується у позиції перед огубленими голосними звуками /O/, /u/, на початку слова перед приголосним, в середині та в кінці слова після голосного:

w	wowk	wowk	volf	вовк
---	------	------	------	------

Примітки:

- Так як реалізація такого фонологічного ефекту як регресивна палаталізація приголосних в середині слів залежить від ступеню м'якості (повної чи напів-) наступного приголосного та здібності попереднього звука бути пом'якшеним (повністю чи неповністю), то вбачається доцільним (як, наприклад, при написанні відповідних

алгоритмів для технологій текст-на-озвучення, а саме, конверсії графема-у-фонему) мати в наявності IPA окремі символи, що відповідали б за ступінь пом'якшення приголосних.

- Оскільки в машинному алфавіті Samra для позначення палаталізації пропонується вибір з двох знаків, то для української мови символ / ' / відповідає за повну палаталізацію, а / _j / відповідно за напівпалаталізацію.

Додаток:

Тлумачення деяких символів, які входять до розширення **X-SAMPA** і використовуються для української мови.

Приголосні

Р (або v\)	губно-зубний апроксимант
h\	глотковий щілинний, дзвінкий

Суперсегментні компоненти

"	основний наголос
%	другорядний наголос

Діакритики

' (обо _j)	палаталізація	напр. t' (або t_j)
_d	зубний	напр. n_d
^	нескладотворчий	напр. I^



Ukrainian Sampa

Cardinal Vowels

The Standard Ukrainian vowel system comprises 6 cardinal vocalic phonemes, as follows:

SAMPA symbol	Transcription of the keyword	IPA	English gloss	Orthography
A	rAk	rak	crayfish	рак
E	stEp	stɛp	steppe	степ
I	kIt	kɪt	whale	кит
O	sOn	sɔn	dream	сон
u	rux	rux	movement	рух
i	k_j'it	kʲit	cat	кіт

Vocalic Allophones

In unstressed position, after semi-consonant /j/ or after palatalised consonants there is a slight qualitative reduction of vowels /A, u/.

6	"zOn6	'zɔnɐ	zone	зона
U	tU"dI	tʊ'dɪ	there	туди

The preceding unstressed /O/ influenced by the following stressed /u/ becomes /o/ and the preceding unstressed /I/ and /E/ influenced by the following stressed /E/ become /e/, so called vowel harmony.

o	to"mu	to'mu	because	тому
e	me"nE, me'nE	me'nɛ me'nɛ	will pass away, me	мине, мене

Sharp reduction, weakening of articulation of unstressed vowels is not typical for Ukrainian.

Cardinal Consonants

The Ukrainian consonant system comprises 32 cardinal consonantal phonemes. Among them 22 'hard' (non-palatalised) phonemes, where /d, t, s, z, dz, ts, n, l/ - dental:

SAMPA Symbol	Transcription of the keyword	IPA	English gloss	Orthography
b	bIk	bɪk	bull	бик
p	pAr	par	steam	пар

d	dAr	dar	gift	дар
t	tAm	tam	there	там
g	"grAtI	'gratɪ	bars	ґрати
k	kut	kut	angle	кут
h\	h\Ak	ɦak	hook	гак
x	xOr	xɔr	choire	хор
v\	v\Ir	vɪr	whirlpool	вир
f	fAx	fax	profession	фах
z	zAl	zal	hall	зал
s	sum	sum	sadness	сум
Z	ZAx	zax	horror	жах
S	Sum	ʃum	noise	шум
dz	"dzIg6	'dzɪgɐ	top	дзиґа
ts	tsE	tsɛ	this	це
dZ	dZm_jil'	dʒmʲilʲ	bumble-bee	джміль
tS	tSAs	tʃas	time	час
m	mEtS	mɛtʃ	sword	меч
n	nAs	nas	us	нас
r	rAb	rab	slave	раб
l	lAn	lan	field	лан

Plus the semi-consonant:

j	j6k	ʃek	how	як
---	-----	-----	-----	----

There are also 9 'soft' (palatalised) consonants, which are shown by the symbol / ' / (ASCII 39) written after the consonant symbol. Consonants / z', s', dz', ts' / are dental.

d'	"d'6d'ko	'dʲɛdʲkɔ	uncle	дядько
t'	t'Uk	tʲʊk	bale	тюк
z'	z'6t'	zʲɛtʲ	son-in-low	зять
s'	s'6d'	sʲɛdʲ	sit down	сядь
dz'	dz'Ob	dʒʲɔb	beak	дзьоб
ts'	ts'6	tsʲɐ	this	ця
n'	n'Ux	nʲʊx	scent	нюх
l'	l'Uk	lʲʊk	hatch	люк
r '	r'id	rʲid	generation	рід

Consonantal allophones

There are also 13 'semi-soft' (semi-palatalised) consonants, which are considered as allophones of their non-palatalised paired consonants and shown by the symbol /_j / written after the consonant symbol.

In view of the fact that there is no special IPA symbol for the feature of semi-palatalization symbol [ʲ] is used to mark both semi-palatalised and palatalised consonants. This is phonemically unambiguous as there are no consonants classified in pairs as 'non-palatalised' and 'palatalised'. (See as well NOTE)

m_j	m_jid'	mʲidʲ	copper	мідь
b_j	b_jik	bʲik	side	бік
p_j	p_jid	pʲid	under	під
g_j	g_jips	gʲips	gypsum	гіпс
k_j	k_jin'	kʲinʲ	horse	кінь
x_j	x_jid	xʲid	motion	хід
h_j	h_jimn	hʲimn	anthem	гімн
v_j	v_jin	vʲin	he	він
f_j'	f_jin	fʲin	Finn	фінн
Z_j	"Z_jink6	'zʲinke	woman	жінка
S_j	S_jis't'	ʃʲisʲtʲ	six	шість
dZ_j	bdZ_jil	bdʒʲil	bees	бджіл
tS_j	tS_jip	tʃʲip	plug	чіп

Other allophones

Non-syllabic allophone / I_^ / is categorized as either a non-syllabic **vocalic** allophone of phoneme / i / or a non-syllabic **consonantal** allophone of phoneme / j /:

I_^	h\AI_^	ɦaɪ̯	grove	гай
-----	--------	------	-------	-----

The allophone /w/ of the consonantal phoneme / v_ / appears before rounded vocalic phonemes /O/, /u/, at the beginning of the word before the consonant, in the middle of the word after the vowel and at the end of the word after vowel:

w	wowk	wowk	volf	вовк
---	------	------	------	------

Notes:

- Realization of such a phonological effect in Ukrainian as regressive palatalization of consonants within words depends on the degree of palatalization (complete or semi) of the following participating consonant and the potential to be fully or partially palatalized of the preceding one.

That is why in this case it would be reasonable to have in IPA two separate symbols stand for palatalization and semi-palatalization. In Ukrainian Sampa symbol / ' / stands for complete palatalization and /_j/ for semi-palatalization.

Appendix:

Description of some symbols from **X-SAMPA** (extended SAMPA) used for Ukrainian

Consonants (pulmonic)

P (or v\)	labiodental approximant
h\	glottal fricative, voiced

Suprasegmentals

"	primary stress
%	secondary stress

Diacritics

' (or _j)	palatalized	e.g. t' (or t_j)
_d	dental	e.g. n_d
^	non-syllabic	e.g. I^

Author of the table "Ukrainian Sampa" both in Ukrainian and in English - **Tonia Bilous**

You can kindly contact me by e-mail:

tonia.bilous@skynet.be

ukraspeech@skynet.be