

## Language specific peculiarities Document for

### Tagalog as Spoken in the Philippines

#### 1. Dialects

Tagalog is spoken by around 25 million<sup>1</sup> people as a first language and approximately 60 million as a second language. It is the local language spoken by most Filipinos and serves as a lingua franca in most parts of the Philippines. It is also the *de facto* national language of the Philippines together with English. Although officially the name of the national language is "Filipino", which has Tagalog as its basis, the majority of Filipinos are unaccustomed, even unaware, of this naming convention. Practically speaking, the name "Tagalog" is used by many Filipinos when referring to the national language. "Filipino", on the other hand, is rather ambiguous and, it must be said, carries nationalistic overtones<sup>2</sup>.

The population of Tagalog native speakers is primarily concentrated in a region that covers southern Luzon and nearby islands. In this project, three major dialectal regions of Tagalog will be presented:

Dialectal Region	Description
North	Includes the Tagalog spoken in Bulacan, Bataan, Nueva Ecija and parts of Zambales.
Central	The Tagalog spoken in Manila, parts of Rizal and Laguna. This is the Tagalog dialect that is understood by the majority of the population.
South	Characteristic of the Tagalog spoken in Batangas, parts of Cavite, major parts of Quezon province and the coastal regions of the island of Mindoro.

#### 1.1. English Influence

The standard Tagalog spoken in Manila, the capital of the Philippines, is also considered the prestige dialect of Tagalog. It is the Tagalog used in the media and is taught in schools. From a lexical perspective, it has absorbed, and continues to absorb, a large number of English words and phrases. The use of English borrowings and code-switching has become so rampant that it has become a defining characteristic of standard Manila Tagalog. While this Tagalog-English hybrid is commonly referred to as "Taglish", it is very difficult to distinguish the language that is acceptable Tagalog from Taglish. The fact remains that some English borrowings have become part of modern Tagalog. In some cases, it is not only ideal to use English but also more natural. Take for example the phrase:

<sup>1</sup> www.ethnologue.com

<sup>2</sup> In the 1930s, Filipino nationalists believed the creation of a national language one of the key elements of nation building. Leaders and language experts from various indigenous language groups were assembled to discuss matters concerning the national language. This resulted in the promotion of "Pilipino", which was based on Tagalog, as the national language. Decades later, this was changed to its present name - "Filipino".

*Mag-sho-shopping kami pagkatapos.* [We will go shopping afterwards]<sup>3</sup>

To express this in pure Tagalog sounds unnatural and rather lengthy:

*Mamimili kami ng mga bagay-bagay pagkatapos.* [We will buy things afterwards]

Similarly, the phrase:

*Mag-sho-shopping kami sa mall pagkatapos para sa birthday ni daddy.* [We will go shopping at the mall afterwards for dad's birthday]

can still be considered Tagalog despite the inclusion of more English borrowings.

The influence of English in Tagalog is substantial and this is reflected in the speech of Tagalog speakers, particularly those in Manila. It is expected that in any large collection of Tagalog speech data, there will be a significant proportion of English. This data collection project is no exception.

## 1.2. Special Handling of Dialects

In this project, three major dialectal regions of Tagalog will be presented.

## 2. Deviation from Native-Speaker Principle

Speech data will be collected mainly from native speakers. In addition, multilingual Filipino speakers, who are fluent in Tagalog and one or more local languages, will also be considered.

## 3. Special Handling of Spelling

There have been a number of spelling systems adopted for the Tagalog language. Although all of the Tagalog spelling systems were based on the Roman script, there were considerable differences. This has led to multiple orthographic forms of many Tagalog words and foreign borrowings, especially those from English and Spanish. In this section, a brief history of the Tagalog spelling system will be presented, followed by a general overview of the spelling guidelines that will be used for this project.

The first spelling system, introduced by the Spanish and based on the Spanish spelling system, was quite effective due to the relative consistency in which sounds were represented by letters or letter combinations. It was also very easy to implement since Tagalog had a smaller phoneme inventory compared to Spanish and there was little need to introduce additional spelling rules. Although, this system is no longer in use, its effects can still be observed in many Philippine names. In fact, the majority of place names in the Philippines are spelled using the early Spanish system. For instance, the names of the Tagalog provinces "Bulacan" (or "Bulacán") and "Batangas" use the Spanish system. If these names were written using the new Tagalog system, they would be spelled as "Bulakan" and "Batanggas" respectively.

In the middle of the last century, a spelling reform was proposed. The new spelling system utilized a 20-letter alphabet - "A B K D E G H I L M N NG O P R S T U W Y" - and promoted one simple rule --- spell the word, as you say it. The Balarila system (a shortened form of the phrase "*bahala ang dila*" [let the tongue decide] was very successful mainly because it was a system that was easy to learn and to teach to a new generation of Filipinos. In this system, foreign borrowings were nativized before being adopted

---

<sup>3</sup> In the document, English Translations of Tagalog terms are contained in square brackets. The translation may not be a literal translation but rather a more natural translation in the target language.

into the written language. For example, from Spanish, the word "coche" became "kotse" [car], "cinco" became "singko" [five], "vaso" - "baso" [glass], "familia" - "pamilya" [family]. The English word "teacher" became "titser", "jeep" became "dyip" (sometimes "diyip"), "traffic" - "trapik", "school" - "iskul". For a while, this system was the standard way in which Tagalog was written.

Due the emergence of globalization in the latter part of the last century and a very successful bilingual educational system, this "Balarila" system was under a lot of strain. English words were being absorbed in the language by the hundreds, if not by the thousands. The process of nativization of English words was becoming very cumbersome. Many were less willing to use the "Balarila" system for foreign loan words and, instead, retained the original spelling of the foreign word. Eventually, in the 1980s, the original alphabet was expanded to include eight additional letters - C, F, J, Ñ, Q, V, X, Z. However, spelling rules were still ambiguous especially on the issue of foreign words. As late as 2006, special forums sponsored by the Commission on the Filipino Language were still being held on the standardization of Filipino (Tagalog) spelling. That same year, spelling guidelines were approved and agreed upon by the delegates, many of whom were language specialists. Still, the new guidelines are not widespread nor widely accepted. In fact, many publishing houses and individual writers maintain their own personalized system of writing Tagalog.

### 3.1. Spelling Guidelines for Tagalog

Due to the inconsistencies in Tagalog spelling, particularly in foreign borrowings, it is essential to implement certain spelling guidelines in order to achieve a high level of consistency. Here are the proposed spelling guidelines that will be used for this project. It should be noted that these guidelines do not address all possible instances and the spelling of some words may need to be considered on a case-by-case basis.

- 1.) Proper nouns retain their original spelling.
- 2.) Tagalog and Spanish words are spelled using the traditional "Balarila" system.
- 3.) For foreign borrowings from English, the original spelling is retained for consistency, except in a few words which have been fully absorbed into the language. Examples of such words, some of which have taken a different semantic meaning from the English original, are: "pulis" [police], "isnab" [snob], "istambay"<sup>4</sup> [standby] and "gimik"<sup>5</sup> [gimmick].

3.1) Prefixation with English borrowings. When a Tagalog prefix is attached to an English borrowing, the prefix is separated by a hyphen ("-") from the loan word.

*nag-shopping* [to have gone shopping]  
*naka-long-sleeves* [to be wearing a long-sleeved shirt]  
*pa-try* [to let (someone) try]

3.2) Partial Reduplication (PR) of English borrowings. When an English borrowing undergoes partial reduplication, the original spelling of the duplicated part of the word is retained and separated from the English root using a hyphen ("-").

mag-shopping + <PR> => *mag-sho-shopping* [will go shopping]  
agenda + <PR> => *age-agenda* [agendas]  
i-forward + <PR> => *i-fo-forward* [will forward (something)]

---

<sup>4</sup> An "istambay", commonly shortened to "tambay", is a person who is generally unemployed and stereotypically hangs around ("stands by") street corners, shops all day long to pass the time.

<sup>5</sup> In addition to the English meaning "publicity stunt", the Tagalog "gimik" is also a general term for a social activity - "Anong gimik ninyo mamayang gabi?" ["What are your plans for tonight", literally, "what is your 'gimik' tonight?"]

3.3) Infixation with English borrowings. When the infixes "in" and "um" are used in English borrowings, a change in the original spelling can be made to avoid confusion in the pronunciation.

forward + "in"-infix => *finorward* [to have forwarded (something)]  
graduate + "um"-infix => *gumraduate* [to have graduated]<sup>6</sup>  
compose + "in"-infix => *kinompose* [to have composed (something)]

In the last example above, the "c" of the English word "compose" is changed to "k" to aid in the proper pronunciation of the initial letter.

In cases where infixation follows partial reduplication of an English borrowing, a hyphen is used to separate the English root from the resulting morpheme.

forward + <PR> + "in"-infix => *fino-forward* [to be forwarding (something)]  
graduate + <PR> + "um"-infix => *guma-graduate* [to be graduating]  
compose + <PR> + "in"-infix => *kino-compose* [to be composing (something)]

---

<sup>6</sup>Note that the alternative form "grumaduate" is also possible depending on the speaker.

#### 4. Description of Character Set Used for Orthographic Transcription

Tagalog uses the same character set as English with the addition of one letter, "ñ", which is taken from Spanish. The Unicode range for printable characters in Tagalog is from U+0020 to U+007E. The letter "ñ" has two codes: U+00D1 (uppercase) and U+00F1 (lowercase). The character map for Tagalog is presented below:

UNICODE	CHARACTER	UNICODE	CHARACTER	UNICODE	CHARACTER
U+0020	[space]	U+0040	@	U+0060	`
U+0021	!	U+0041	A	U+0061	a
U+0022	"	U+0042	B	U+0062	b
U+0023	#	U+0043	C	U+0063	c
U+0024	\$	U+0044	D	U+0064	d
U+0025	%	U+0045	E	U+0065	e
U+0026	&	U+0046	F	U+0066	f
U+0027	'	U+0047	G	U+0067	g
U+0028	(	U+0048	H	U+0068	h
U+0029	)	U+0049	I	U+0069	i
U+002A	*	U+004A	J	U+006A	j
U+002B	+	U+004B	K	U+006B	k
U+002C	,	U+004C	L	U+006C	l
U+002D	-	U+004D	M	U+006D	m
U+002E	.	U+004E	N	U+006E	n
U+002F	/	U+004F	O	U+006F	o
U+0030	0	U+0050	P	U+0070	p
U+0031	1	U+0051	Q	U+0071	q
U+0032	2	U+0052	R	U+0072	r
U+0033	3	U+0053	S	U+0073	s
U+0034	4	U+0054	T	U+0074	t
U+0035	5	U+0055	U	U+0075	u
U+0036	6	U+0056	V	U+0076	v
U+0037	7	U+0057	W	U+0077	w
U+0038	8	U+0058	X	U+0078	x
U+0039	9	U+0059	Y	U+0079	y
U+003A	:	U+005A	Z	U+007A	z
U+003B	;	U+005B	[	U+007B	{
U+003C	<	U+005C	\	U+007C	
U+003D	=	U+005D	]	U+007D	}
U+003E	>	U+005E	^	U+007E	~
U+003F	?	U+005F	_	U+00D1	Ñ
				U+00F1	ñ

## 5. Description of Romanization Scheme

A Romanization scheme is not needed for standard Tagalog.

## 6. Description of Method for Word Boundary Detection

Word boundaries in the orthography are determined by localization of white spaces (blank, tab, etc).

## 7. Table containing all phonemes in the stipulated notation

IPA	SAMPA	Example Word	Transcription
<b>CONSONANTS</b>			
p	p	puno	" p u . n o ?
b	b	bato	b 6 . " t o
t	t	tatay	" t a . t a j
d	d	dila	" d i . l a ?
k	k	kabayo	k 6 . " b a . j o
g	g	gabi	g 6 . " b i
ʔ	ʔ	isda	? i s . " d a ?
tʃ	tʃ	kotse	" k o . tʃ e
dʒ	dʒ	jeepney	" dʒ i p . n i
s	s	sulok	" s u . l o k
ʃ	ʃ	siyam	" ʃ a m
h	h	kahon	k 6 . " h o n
m	m	mesa	m u n . " d o
n	n	nayon	" n a . j o n
ŋ	ŋ	ngipin	" ŋ i . p i n
l	l	luto	" l u . t o ?
r	r	relo	r e . " l o
j	j	yaman	" j a . m a n
w	w	walis	w 6 . " l i s
<b>VOWELS</b>			
i	i	ilaw	" ? i . l a w
e	e	Quezon	" k e . s o n
ɐ	ɐ	ano	? ɐ . " n o
a	a	daan	d 6 . " ? a n

IPA	SAMPA	Example Word	Transcription
o	o	oso	" ? o . s o
u	u	ulit	" ? u . l i t
aɪ	aj	hintay	h i n . " t a j
eɪ	ɛj	bahay-bahayan	b ɛ . % h ɛ j . b ɛ . " h a . j a n
eɪ	ej	reyna	" r e j . n a
oɪ	oj	kasoy	k ɛ . " s o j
uɪ	uj	tuluy-tuloy	t u . % l u j . t u . " l o j
aʊ	aw	palitaw	p a . l i . " t a w
eʊ	ɛw	araw-arawin	? ɛ . % r ɛ w . ? ɛ . " r a . w i n
iʊ	iw	giliw	" g i . l i w
oʊ	ow	show	" S o w
SUPRASEGMENTALS			
'	"	primary stress	
'	%	secondary stress	
.	.	syllable break	

## 8. Complete List of all rare phonemes

### 8.1. List of Foreign phonemes

FOREIGN PHONEMES			
f	f	Fernando	f e r . " n a n . d o
v	v	video	" v I d . j o w
z	z	zoo	" z u
ʒ	Z	vision	" v I . Z @ n
θ	T	thing	" T I N
ð	D	this	" D I s
ɪ	I	sit	" s I t
ɛ	E	bet	" b E t
æ	{	cat	" k { t
ɜ	3	first	" f 3 r s t
ə	@	above	? @ . " b V v
ɑ	A	hot	" h A t
ʌ	V	cut	" k V t
ʊ	U	book	" b U k

The foreign phonemes presented above are very rare and are used mainly by those who are well educated and live in high-density cosmopolitan areas. All these phonemes come from English and in general, words of English origin, will have multiple pronunciations due to considerable differences between the native Tagalog phoneme inventory and that of English. The inclusion of these foreign phonemes allows for a more accurate description of the pronunciation of speakers of contemporary Tagalog.

## 9. Other Language Specific Items

### 9.1. Spelling Alphabet

Below is the modern Tagalog alphabet and the corresponding SAMPA pronunciation of each letter.

Letter	Pronunciation
A	" e j
B	" b i
C	" s i
D	" d i
E	" i
F	" e f
G	" dZ i
H	" e j tS
I	" a j
J	" dZ e j
K	" k e j
L	" e l
M	" e m
N	" e n
Ñ	" e n . j e
NG	" e n # " dZ i
O	" o w
P	" p i
Q	" k j u
R	" a r
S	" e s
T	" t i
U	" j u
V	" v i
W	" d a . b 6 l . j u
X	" e k s
Y	" w a j
Z	" z i



## 9.2. Table of Numbers

Number	Tagalog	English	Nativized Spanish
0	zero	zero	sero
1	isa	one	uno
2	dalawa	two	dos
3	tatlo	three	tres
4	apat	four	kuwatro
5	lima	five	singko
6	anim	six	sais
7	pito	seven	siyete
8	walo	eight	otso
9	siyam	nine	nuwebe
10	sampu	ten	diyes
11	labing-isa	eleven	onse
12	labindalawa	twelve	dose
13	labintatlo	thirteen	trese
14	labing-apat	fourteen	katorse
15	labinlima	fifteen	kinse
16	labing-anim	sixteen	disisais
17	labimpito	seventeen	disisiyete
18	labingwalo	eighteen	disiotso
19	labinsiyam	nineteen	disinuwebe
20	dalawampu	twenty	beynte, bente <sup>7</sup>
21	dalawampu't isa	twenty-one	beynti-uno
22	dalawampu't dalawa	twenty-two	beyntidos
30	tatlumpu	thirty	treynta, trenta <sup>7</sup>
31	tatlumpu't isa	thirty-one	treynta y uno
40	apatnapu	forty	kuwarenta
50	limampu	fifty	singkuwenta
60	animnapu	sixty	sisenta
70	pitumpu	seventy	sitenta
80	walumpu	eighty	otsenta
90	siyamnapu	ninety	nobenta
100	isang daan	one hundred	siyento
200	dalawang daan	two hundred	dos siyento
1000	isang libo	one thousand	mil
10000	sampung libo	ten thousand	diyes mil
1,000,000	isang milyon	one million	(not in common use)
10,000,000	sampung milyon	ten million	(not in common use)

Tagalog speakers use a unique trilingual number system in expressions containing numbers. This system varies from speaker to speaker based on his/her educational and social background. Despite these variations, there are some general trends. Here are some of them:

<sup>7</sup> The spelling variants "bente" and "trenta" reflect a common colloquial pronunciation.

- Tagalog is more consistently used in counting objects, people, etc. For example, *dalawang lapis* [two pencils], *tatlong estudyante* [three students], *labindalawang bansa* [twelve countries]. It is also used in expressing small denominations of money: *dalawang piso* [two pesos], *limang piso* [five pesos] etc.
- Spanish numbers are common in:
  - expressions of time (see section on time expressions)
  - specific denominations of money - a twenty-five cent piece is a "*beyntisingko*", a fifty-cent piece, a "*singkuwenta*", a twenty-peso bill is more commonly called a "*beynte*"
  - attributive expressions of age - "*trenta-anyos na aktor*" [a thirty year-old actor], "*dose-anyos na rum*" [a twelve year-old rum]
  - certain set expressions in common use - "*dekada nobenta*" [the nineties], "*trenta porsiyento*" [thirty percent].
- English is used more often than Tagalog or Spanish in numerical expressions especially in the following:
  - Phone numbers
  - Commodity prices - a figure like ₱99.95 is almost without exception expressed in English
  - Banking and Business concepts, e.g., PIN numbers, credit card numbers, market indexes
  - Dates especially when the date is a complete date as in "February 12, 1994"
  - Time expressions especially when the time is non-exact "3:41 pm"
  - House numbers, zip codes, scoring systems in sports, marks in school, and many more.