

Documentation

GLOBALPHONE: Dictionaries in multiple Languages - Spanish -

Version 1.0



December 18, 2006

**© XLingual, LLC
5529 Kentucky Avenue, 2nd floor, Pittsburgh, PA 15232**

Copyrights:

This document describes the formats and phone sets of the **GlobalPhone dictionaries** for multiple languages. All use of this corpus is subject to a license agreement. Copyright holder of the data is XLingual, LLC. No parts of these dictionaries shall be distributed or reproduced in any form or by any means without permission from the copyright holder.

I. Dictionary Format

A pronunciation dictionary of a particular language contains the pronunciations of all word forms found in the transcription data of the GlobalPhone audio recordings of this language. The word forms are given in the romanized script as explained in section „Romanization“ for each language. The format of the dictionary is very straight forward. Each line consists of one word form and its pronunciation. Words forms and their pronunciations are separated by tabulator. The pronunciation consists of a concatenation of phone symbols separated by blanks. Pronunciation variants are indicated by (<n>) with n = 2, 3, 4, ... showing the number of variants per word. The order in which variants occur is not necessarily related to their frequency in the corpus. The set of phone symbols used is explained in section „Phone set“ for each language. Below is a short excerpt from the Spanish dictionary.

espan5ol	e s p a n~ o l
espan5ola	e s p a n~ o l a
espanta	e s p a n t a
espantosa	e s p a n t o s a
esparadrapo	e s p a r f a D r f a p o
esparcidas	e s p a r f s i D a s
esparcidas(2)	e s p a r f T i D a s

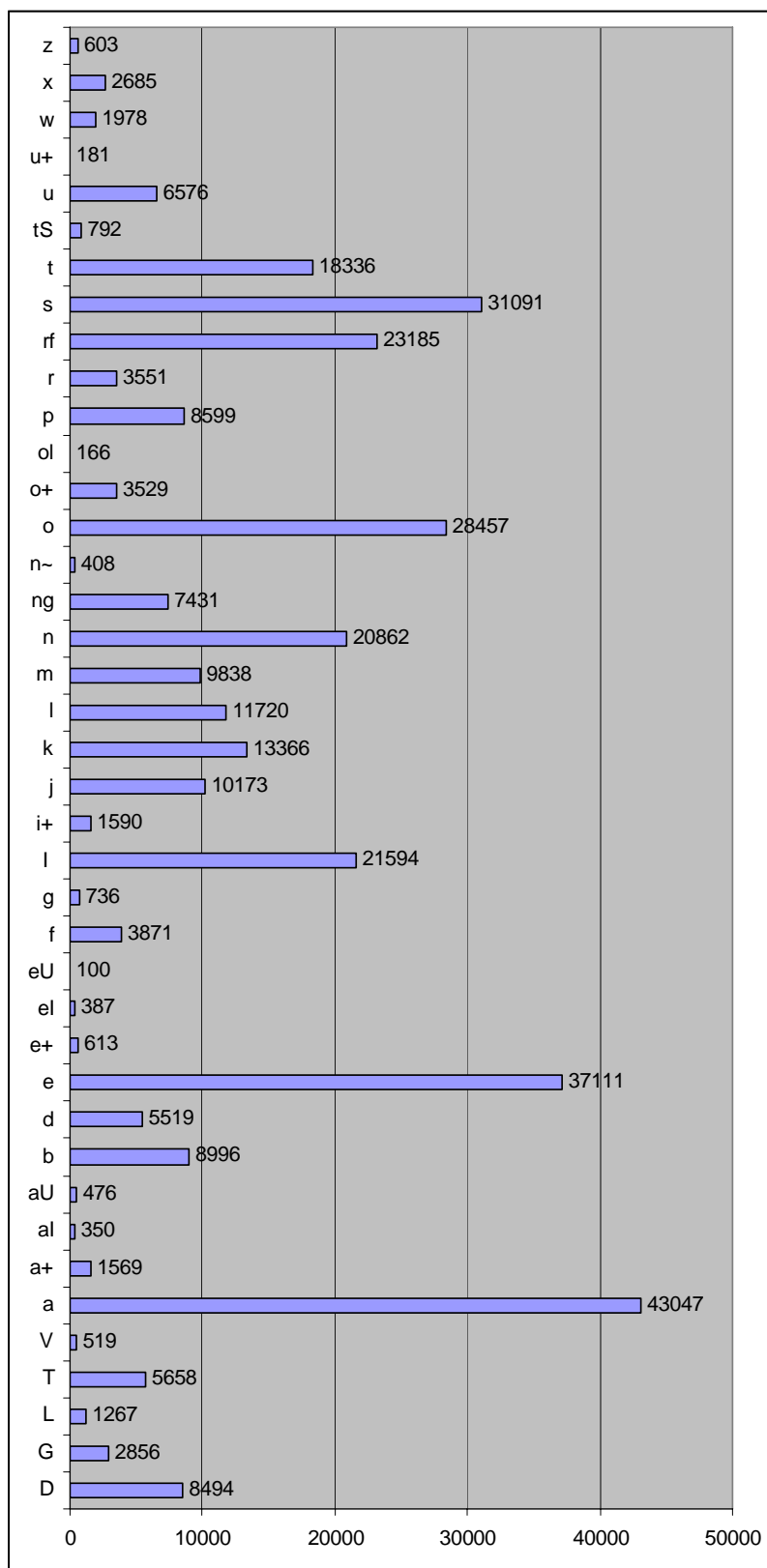
II. Romanization

The following list describes the original characters/diacritics used in Spanish script, the corresponding ISO8859-1 code and romanized form as used in the romanized transcription files and the pronunciation dictionary.

Original Character	ISO-code	Romanized character
á	\341	a+
Á	\301	A+
é	\351	e+
É	\311	E+
í	\355	i+
Í	\315	I+
ó	\363	o+
Ó	\323	O+
ú	\372	u+
Ú	\332	U+
ü	\374	u^
Ü	\334	U^
ñ	\361	n5
Ñ	\321	N5

III. Phone Set

The phone set as used in the Spanish pronunciation dictionary is illustrated below using the International Phonetic Alphabet IPA. The red semi-transparent boxes in the IPA table mark those phones, which occur in the language. The red bold letters give the names of the phones as used in the pronunciation dictionary. In total the phone set consists of 40 phones, 25 consonants (D, G, L, T, V, b, d, f, g, j, k, l, m, n, ng, n~, p, r, rf, s, t, tS, w, x, z) and 15 vowels (a, a+, aI, aU, e, e+, eI, eU, I, i+, o, o+, oI, u, u+). These names correspond to the GlobalPhone notation M_, i.e. D=M_D, G=M_G, and so on.



In total the dictionary contains 41283 entries from 31993 different word forms occurring in the transcriptions of the Spanish part of the GlobalPhone data collection. 9290 pronunciation variants are given. The figure above shows the frequencies of occurrence of the 40 phones in this dictionary.

Spanish Consonants

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

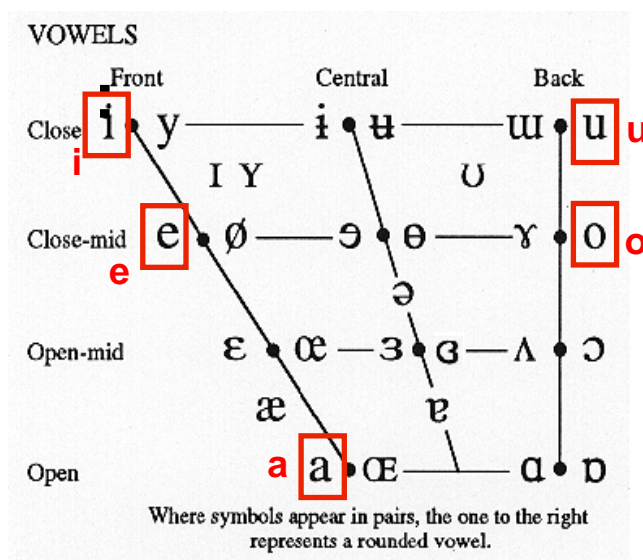
CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m			n		ɳ	ɲ	ŋ			
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j		ɰ		
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

tʃ Affricate t s

Spanish Vowels



Stressed Vowels

i+ stressed i
e+ stressed e
a+ stressed a
o+ stressed o
u+ stressed u

Diphthongs

oi
eU
el
aU
al