

Documentation

GLOBALPHONE: Dictionaries in multiple Languages - FRENCH -

Version 1.0



July 23, 2005

**© 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

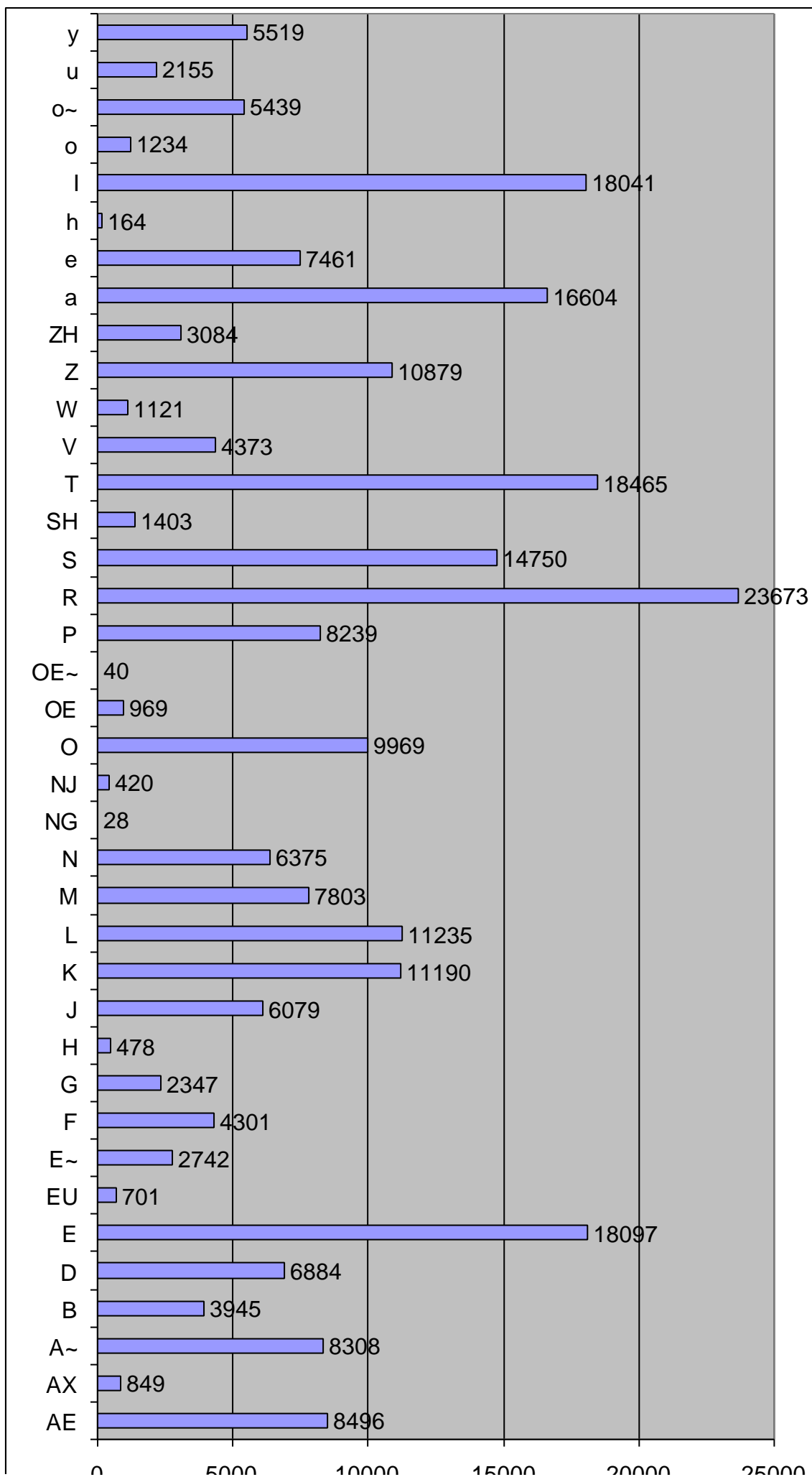
A pronunciation dictionary of a particular language contains the pronunciations of all word forms found in the transcription data of the GlobalPhone audio recordings of this language. The word forms for French are given in the original script. The format of the dictionary is very straight forward. Each 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 $n = 2, 3, 4, \dots$ 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 French dictionary.

framboises	F R A~ B W a Z
framboises(2)	F R A~ B W a Z AE
franc	F R A~
franc(2)	F R A~ K
france	F R A~ S
france(2)	F R A~ S AE
franche	F R A~ SH

II. Phone Set

The phone set as used in the French 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 38 phones, 22 consonants (P, B, T, D, K, G, M, N, NJ, NG, F, V, S, Z, SH, ZH, R, h, L, J, W, H), and 16 vowels (i, y, u, e, EU, o, E, OE, AX, O, a, AE, A~, E~, o~, OE~).

In total the dictionary contains 36837 entries from 20710 word forms occurring in the transcriptions of the French part of the GlobalPhone data collection. The figure below shows the frequencies of occurrence of the 38 phones in this dictionary.



French Consonants

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

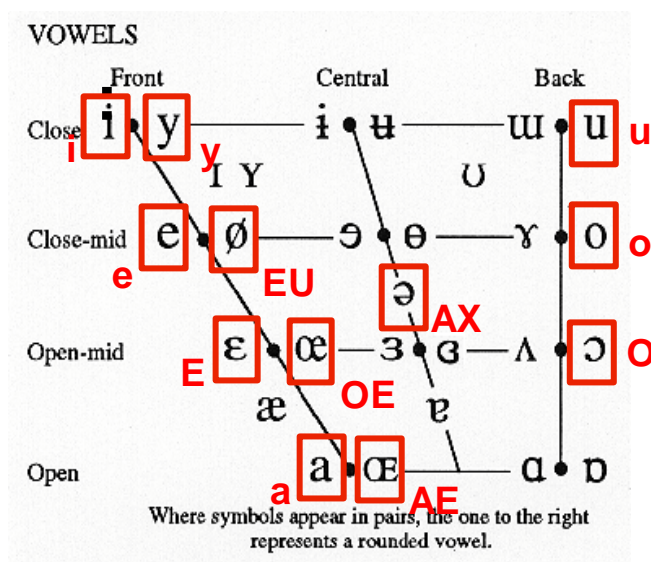
CONSONANTS (PULMONIC)

	P	B		Dental	T	D		Retroflex		K	G			
	Labial	Labiodental			Alveolar	Postalveolar			Palatal	Velar		Uvular	Pharyngeal	Glottal
Plosive	p	b			t	d		ʈ ɖ	c ɟ	k	g	q ɢ		ʔ
Nasal		m	M	ɱ		n	N	ɳ		ɲ	NJ	ŋ	N	NG
Trill		B				r						R		
Tap or Flap						ɾ		ɽ						
Fricative	ɸ β		F	V		s	z	ʃ ʒ	ç ʝ	x	χ	R	ħ ʕ	h
		f	v			θ ð		ʂ ʐ				ʁ		
Lateral fricative						ɬ ɮ								
Approximant			ʋ			ɭ		ɻ		j	w			
						ɮ								
Lateral approximant						ɭ		ɻ						

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

Approximant **H** **H** (labial-velar)

French Vowels



Where symbols appear in pairs, the one to the right represents a rounded vowel.

Nasal Vowels

in **E~** nasal E

an **A~** nasal A, a

on 0~ nasal o

un OE ~ nasal OE