phonItalia: a phonological lexicon for Italian

Version 1.01 - 9th July 2013

Document Title: Descriptions of database fields and phonological alphabet

phonItalia Word Forms Fields

NumLetters	Number of letters in the word
OrthVCV	The consonant vowel structure of the orthographic representation of the word
OrthUniq	Orthographic uniqueness point
OrthUniqM1	Orthographic uniqueness point minus one
NumHomographs	Number of homographs
Phones	The phonological representation of the word form
PhonSyll	Phonological representation of the word form with syllable boundaries (.)
NumPhones	Number of phones in the word
PhonVCV	Consonant vowel structure of the phonological representation of the word
NumSylls	Number of syllables in the word
PhonUniq	Phonological uniqueness point
PhonUniqM1	Phonological uniqueness point minus one
NumHomophones	Number of homophones
Orth_N	Size of the orthographic neighbourhood
Orth_N_MFreq	Mean log frequency of the orthographic neighbourhood
Orth_N_G	Number of orthographic neighbours with a higher frequency than the word
Orth_N_L	Number of orthographic neighbours with a lower frequency than the word
Orth_N_G_MFreq	Mean log frequency of the orthographic neighbours with a lower

	frequency than the word
Orth_N_L_MFreq	Mean log frequency of the orthographic neighbours with a higher frequency than the word
Orth_N_RelFreq	Relative log frequency of the current word and that of its orthographic neighbourhood
OLD	Orthographic Levenshtein Distance 20
OLDF	Mean log frequency of words of the 20 words used to calculate the OLD
OLD_RelF	Relative log frequency of the word and the 20 used to calculate the OLD
Phon_N	Size of the phonological neighbourhood
Phon_N_MFreq	Mean log frequency of the phonological neighbourhood
Phon_N_G	Number of phonological neighbours with a higher frequency than the word
Phon_N_L	Number of phonological neighbours with a lower frequency than the word
Phon_N_G_MFreq	Mean log frequency of the phonological neighbours with a lower frequency than the word
Phon_N_L_MFreq	Mean log frequency of the phonological neighbours with a higher frequency than the word
Phon_N_RelFreq	Relative log frequency of the current word and that of its phonological neighbourhood
PLD	Phonological Levenshtein Distance 20
PLDF	Mean log frequency of words of the 20 words used to calculate the PLD
PLD_RelF	Relative log frequency of the word and the 20 used to calculate the PLD

Fields inherited from Colfis[1] database

[1](Laudanna, Thorton, Brown, Burani, & Marconi, 1995; Bertinetto et al., 2005)

nLem	The associated lemma index number assigned to each of the Colfis word forms. This number can be used to match the wordform with the lemma in the Colfis lemma database.
fqTot	Total absolute frequency of the word form.
fqTotL	Total log frequency of the word form [fqTot]
fqQuo	Absolute frequency from newspapers
fqPer	Absolute frequency from periodical magazines
fqLib	Absolute frequency from books
dispT	Dispersion of total frequency [fqTot]
dispQuo	Dispersion of frequency from newspapers [fqQuo]
dispPer	Dispersion of frequency from periodical magazines [fqPer]
dispLib	Dispersion of frequency from books [fqLib]
fqRelT	Relative total frequency
fqRelQ	Relative frequency from newspapers
fqRelP	Relative frequency from periodical magazines
fqRelL	Relative frequency from books
rango	Word form index number from Colfis.
lung	Number of characters in orthographic word form [word] (excluding ')
word	Orthographic word form
lemma	Orthographic representation of lemma associated with the word form
gramCat	Grammatical category with the following classifications: **B** Adverb **C** Conjunction **E** Noun **G** Adjective **I** Interjection **N** Pronoun **P** Preposition **K** Punctuation **R** Article **S** Substantive

V Verb

X Not identified

Z Symbol

NU Numeral

TC Composed verb

VA Auxilliary verb

U unknown

@ syntagmatic word (used in combination with another code, for example S IN E@, would be a noun in a syntagmatic word)

Derived Sub-Lexical Statistics

Phone Statistic Fields

Single Onset	Statistics for phones found in a single consonant syllable onset. For example, the phone /n/ in the word /a.E.ro.pla.**n**o/.
Onset /Cc/	phones found in the first consonant of a double consonant syllable onset. For example, /p/ in /a.E.ro.**p**la.no/.
Onset /cC/	phones in the second consonant of a double consonant syllable onset. For example, /l/ in /a.E.ro.p**l**a.no/.
Onset /Ccc/	phones in the first consonant of a triple consonant syllable onset. For example, /G/ in /Gan.**G**ljo /.
Onset /cCc/	phones in the second consonant of a triple consonant syllable onset. For example, /l/ in /Gan.G**l**jo/.
Onset /ccC/	phones in the third consonant of a triple consonant syllable onset. For example, /j/ in /Gan.Gl**j**o/.
Nucleus	phones that form the nucleus of a syllable. For example /o/ is twice found as a nucelus in /a.E.r**o**.pla.n**o**/.
Single Coda	provides statistics for phones found in a single consonant syllable coda. For example, /n/ in the word /la**n**.ce/.
1st Coda	phones in the first consonant of a syllable coda (greater than one consonant in length). For example, /l/ in /fi**l**m/.
2nd Coda	phones in the second consonant of a syllable coda (greater than one consonant in length). For example, /m/ in /fil**m**/.
Geminate	phones that are found in geminate position in a word. For example, /g/ in the word /ma**g**.**g**o.re/.

Syllable Statistic Fields

MonoSyll	provides frequency information for syllables that occur in monosyllabic words
Initial	describes syllables that occur word initially in multisyllabic words, for example /ti/ in /**ti**.fa.no/.
Medial	syllables found in multisyllabic words that are neither word initial nor final, for example /ti/ in /ul.**ti**.mo/.
Final	syllables found in multisyllabic words that are word final, for example, /ti/ in /van.**ti**/.
PhonSeq_Total	frequency of occurrence for the phone sequence of the syllable in the lexicon irrespective of syllable boundaries. e.g. /par/ in the word /pre.**pa.r**a/
PhonSeq_Word_Initial	Only includes the statistics for words where the syllable phone sequence is found word initially. For the syllable /tar/ it would include an occurrence for the word /**ta.r**a.re/, but not in /kon.**ta.r**e/.

Syllable Onset and Coda Fields

Total	statistics for syllable onsets or codas found in any word position
Word Initial	statistics for syllable onsets found in word initial position, for example, /t/ in /**t**i.fa.no/
Word Medial	syllable onsets and codas that are medial to the word. For example, the onset /d/ or the coda /n/ in /mon.**d**o/
Word Final	only syllable codas that are found in word final position
Geminate	subset of the word medial statistics, and is limited to syllable onsets or codas that are geminate, for example, the onset and coda /l/ in /al.lo/.
Number of phones	number of phones in the syllable onset or coda
1st phone	1st (leftmost) phone in the syllable onset or coda, for example /p/ in the onset /**p**l/, or /l/ in the coda /**l**m/.
2nd phone	2nd phone in syllable onset or coda, for example /l/ in the onset /p**l**/, or /m/ in the coda /l**m**/.
3rd phone	3rd phone in syllable onset or coda, this would be blank in the example of /pl/, or would be /s/ in the coda /rk**s**/ from 'Marx'.
4th phone	4th phone in syllable onset (this field is missing in the coda database)

Character-bigram and Biphone Fields

Word Initial	statistics of bigrams that occur in word initial position. For example, the biphone /ko/ in /kon.trad.det.te/ or the character bigram 'se' in 'sempre'.
Word Medial	bigrams that occur word medially, For example, the biphone /on/ in /kon.trad.det.te/ or the character bigram 'mp' in 'sempre'.
Word Final	bigrams that occur word finally. For example, the biphone /te/ in /kon.trad.det.te/ or the character bigram 're' in 'sempre'.
Syllable Onset	biphones that are found in syllable initial position, for example /tr/ in /kon.trad.det.te/. This would include all occurrences in which the first and second phone of the biphone and syllable were shared.
Syllable Medial	biphones found in syllable medial position, for example /ra/ in /kon.trad.det.te/. This would include all occurrences where neither the first or second phone of the biphone coincided with the initial or final phone of a syllable.
Syllable Final	biphones that are found in syllable final position, for example /et/ in /kon.trad.det.te/. This would include all occurrences in which the final and penultimate phone of the bigram and a syllable were shared.
Cross Syllable	biphones that cross syllable boundaries. For example, /nt/ in /kon.trad.det.te/. In this case the first phone of the biphone must consist of the final phone of the syllable preceding the boundary, and the second phone the first phone of the syllable that proceeds the boundary.

Description of the phonological alphabet used in phonItalia

Phone (IPA)	Phone (ascii)	Example (orthographic)	Example (phonological)
a	a	R**a**ta	/r**a**ta/
i	i	M**i**te	/m**i**te/
o	o	D**o**ve	/d**o**ve/
e	e	R**e**te	/r**e**te/
u	u	M**u**to	/m**u**to/
ɛ	E	M**e**ta	/m**E**ta/
ɔ	O	**M**oto	/m**O**to/
t	t	**T**ana	/**t**ana/
r	r	**r**ete	/**r**ete/
n	n	**n**occa	/**n**Okka/
s/z	s	**s**ano	/**s**ano/
l	l	**l**ama	/**l**ama/
k	k	**C**ane	/**k**ane/
m	m	m**o**lla	/**m**Olla/
p	p	**P**ane	/**p**ane/
d	d	**D**anno	/**d**anno/
v	v	**v**ano	/**v**ano/
j	j	**i**eri	/**j**Eri/
b	b	**B**anco	/**b**anko/
f	f	**f**ame	/**f**ame/
tʃ	c	**c**ena	/**c**ena/
ts	z	**z**itto	/**z**itto/
ʤ	g	**g**amba	/**g**amba/
g	G	**g**atto	/**G**atto/
w	w	**u**omo	/**w**Omo/
ʎ	L	**gl**i	/**L**i/
dz	Z	**z**ona	/**Z**Ona/
ʃ	S	**sc**endo	/**S**endo/
ɲ	N	o**gn**i	/o**NN**i/

All enquires, corrections, or requests for further information can be directed to
information@phonitalia.org

phonItalia was developed by Jeremy Goslin[1], Claudia Galuzzi[2], & Cristina Romani[3]

[1]School of Psychology, University of Plymouth,UK.

[2]Fondazione Santa Lucia, i.r.c.s.s., Roma , Italy.

[3]School of Life and Health Sciences, Aston University, UK.