Supplemental Material S12. AFC-list columns: Names and descriptions.

This document describes the column names used in the AFC-list. For each column, a brief description of its content is given.

AFC-list column heading	Description
OrthographicForm	The orthographic form
GIL2P_OT_Average	The average grapheme-initial letter-to-pronunciation orthographic transparency
GIL2P_OT_Min	The minimum grapheme-initial letter-to-pronunciation orthographic transparency
PIP2G_OT_Average	The average pronunciation-initial phone-to-grapheme orthographic transparency
PIP2G_OT_Min	The minimum pronunciation-initial phone-to-grapheme orthographic transparency
G2P_OT_Average	The average grapheme-to-pronunciation orthographic transparency
UpperCase	The proportion of times the word begins with an upper case letter
Homographs	The reduced phonetic transcriptions of homographs
HomographCount	The number of homographs
SpecialCharacter	The existence of special characters in the orthographic form
RawWordTypeFrequency	The total number of occurrences of the spelling in the internet blog corpora.
RawDocumentCount	The total number of different internet blogs in which the word occurs, based on the internet blog corpora.
PhoneticForm	The standard phonetic transcription, according to the convention described in the present article.
TemporarySyllabification	The phonetic transcription re-syllabified by the syllabification tool used when calculating SSPP.
ReducedTranscription	The reduced phonetic transcription
PhonotacticType	The phonotactic type
SSPP_Average	The average normalized stress and syllable based phonotactic probability
SSPP_Min	The minimum normalized stress and syllable based phonotactic probability
PSP_Sum	The summed positional segment probability
PSBP_Sum	The summed position specific bi-phone probability
S_PSP_Average	The average standardized positional segment probability
S_PSBP_Average	The average standardized position specific bi-phone probability

Homophones	All homophone spellings
HomophoneCount	The number of homophones
PNDP	The Zipf-scale weighted phonetic neighborhood density probability
PLD1Transcriptions	The reduced phonetic transcriptions and raw word frequencies (delimited by colons) of phonetic neighbors (delimited by vertical lines), sorted according to word frequency. The first word is the look-up word.
ONDP	The Zipf-scale weighted orthographic neighborhood density probability
OLD1Spellings	The orthographic forms and raw word frequencies (delimited by commas) of orthographic neighbors (delimited by vertical lines), sorted according to word frequency. The first word is the look-up word.
Sonographs	The sonographs
AllPoS	The word-class assignments, and their relative distributions in the internet blog corpora.
AllLemmas	The lemma assignments, and their relative distributions in the internet blog corpora.
NumberOfSenses	The total number of senses of all lemmas
Abbreviation	The abbreviation marking
Acronym	The acronym marking
ForeignWord	The foreign word marking
ZipfValue	The Zipf-scale value
LetterCount	The number of letters in the orthographic form
GraphemeCount	The number of graphemes in the sonograph array
DiGraphCount	The number of two-letter graphemes in the sonograph array
TriGraphCount	The number of three-letter graphemes in the sonograph array
LongGraphemesCount	The number of graphemes longer than three letters in the sonograph array
SyllableCount	The number of syllables in the phonetic transcription
Tone	The pitch accent
MainStressSyllable	The primary stressed syllable (1-based index)
SecondaryStressSyllable	The secondary stressed syllable (1-based index)
PhoneCount	The number of phonetic segments in the phonetic transcription
PLD1WordCount	The number of phonetic neighbors
OLD1WordCount	The number of orthographic neighbors
PossiblePoSCount	The number of possible word classes
PossibleLemmaCount	The number of possible lemmas