**Project/Data Title:** Italian Age of Acquisition Norms for a Large Set of Words (ItAoA)

**Project/Data Description: (200-500 words brief description of the theory/background for the data)**

Age of acquisition (AoA) represents the age at which a word is learned. This measure has been shown to affect performance in a large variety of cognitive tasks (see reviews by Juhasz, 2005; Johnston and Barry, 2006; Brysbaert and Ellis, 2016), with faster reaction times for words learned early in life compared with those learned later.

There are two main approaches to derive AoA data. First, objective AoA measures can be obtained by analysis of children’s production (Chalard et al., 2003; Álvarez and Cuetos, 2007; Lotto et al., 2010; Grigoriev and Oshhepkov, 2013). Within this approach, children (classified by age) are asked to name the picture of common objects and activities. The AoA of a given word is computed as the mean age of the group of children in which at least 75% of them can name the picture correctly. Alternatively, subjective AoA can be obtained by using adult estimates (Barca et al., 2002; Ferrand et al., 2008; Moors et al., 2013). Here, adult participants are asked to provide ratings of AoA on either a Likert scale (Schock et al., 2012; Alonso et al., 2015; Borelli et al., 2018) or directly in years, by indicating the number corresponding to the age they thought they had learned a given word (Stadthagen-Gonzalez and Davis, 2006; Ferrand et al., 2008; Moors et al., 2013). Compared to the use of a Likert scale, this latter method is easier for participants to use and it does not restrict the response range artificially, instead providing more precise information about the words’ AoA (Ghyselinck et al., 2000). It has been shown that the AoA estimates obtained from the two different methods are highly correlated (Morrison et al., 1997; Ghyselinck et al., 2000; Pind et al., 2000; Lotto et al., 2010; see also Brysbaert, 2017; Brysbaert and Biemiller, 2017) and this correlation still remains significant when other variables, such as familiarity, frequency, and phonological length, are controlled for (Bonin et al., 2004).

Only two sets of Italian norms with objective AoA (Rinaldi et al., 2004) and subjective AoA (Borelli et al., 2018) include abstract and concrete words and different word classes (adjective, noun, and verb), but they are limited to a relatively small number of word stimuli (519 and 512 words, respectively). Unfortunately, the lack of overlap between AoA (Dell’Acqua et al., 2000; Barca et al., 2002; Barbarotto et al., 2005; Della Rosa et al., 2010; Borelli et al., 2018) and semantic-affective norms (Zannino et al., 2006; Kremer and Baroni, 2011; Montefinese et al., 2013b, 2014; Fairfield et al., 2017) for Italian words has prevented the direct comparison of different lexical-semantic dimensions to establish the extent to which they overlap or complement each other in word processing. An important motivation of the present study is to extend previous Italian norms by collecting AoA ratings for a much larger range of Italian words for which concreteness and semantic-affective norms are now available thus ensuring greater coverage of words varying along these dimensions.

**Methods Description:** (brief description of how the data was collected)

A total of 507 native Italian speakers were enrolled to participate in an online study (436 females and 81 males; mean age: 20.82 years, SD = 2.22; mean education: 15.16 years, SD = 1.11). We selected 1,957 Italian words from our Italian adaptations of the original ANEW (Montefinese et al., 2014; Fairfield et al., 2017) and from available Italian semantic norms (Zannino et al., 2006; Kremer and Baroni, 2011; Montefinese et al., 2013). The set of stimuli included 76% of nouns, 16% of adjectives, and 8% of verbs. The word stimuli were presented in the same verbal form as the previous Italian norms (e.g., the verbs were presented in the infinitive form) to preserve the consistency with these data collections (Montefinese et al., 2014; Fairfield et al., 2017). Word stimuli were distributed over 20 lists containing 97–98 words each. In order to avoid primacy or recency effects, the order in which words appeared in the list was randomized for each participant separately. All lists were roughly matched for word length, word frequency, number of orthographic neighbors, and mean frequency of orthographic neighbors. For each list, an online form was created using Google modules. Participants were asked to estimate the age (in years) at which they thought they had learned the word, specifying that this information should indicate the age at which, for the first time they understood the word when somebody else used it in their presence, even when they did not use the word themselves. These instructions and the examples provided to the participants closely matched those used in a large number of previous studies (Ghyselinck et al., 2000; Stadthagen-Gonzalez and Davis, 2006; Kuperman et al., 2012; Moors et al., 2013; Łuniewska et al., 2016). The task lasted about 40 min.

**Data Location:** https://osf.io/rzycf/

**Date Published:** 2019-02-13

**Dataset Citation:** Dataset: Montefinese, M., Vinson, D., Vigliocco, G., & Ambrosini, E. (2018, November 26). Italian age of acquisition norms for a large set of words (ItAoA). <https://doi.org/10.17605/OSF.IO/3TRG2>

Article: Montefinese M, Vinson D, Vigliocco G and Ambrosini E (2019) Italian Age of Acquisition Norms for a Large Set of Words (ItAoA). Front. Psychol. 10:278. doi: 10.3389/fpsyg.2019.00278

**Keywords:** age of acquisition, word, lexicon, Italian language, cross-linguistic comparison, subjective rating

**Use License:** CC-By Attribution 4.0 International

**Geographic Description - City/State/Country of Participants:** Italy

Column Metadata: Fill in the chart below for each column of data in the dataset. Please note you can filter out columns that are not useful for this project.

|  |  |  |
| --- | --- | --- |
| Variable Name | Variable Description | Type (numeric, character, logical, etc.) |
| SS\_ID | Participant code | numeric |
| Ita\_Word | Italian word stimuli | character |
| Eng\_Word | English translation of the stimuli | character |
| Rating | Participants’ age of acquisition rating for each word | numeric |

What columns should we use to simulate the data?

* **Item labels are found:** Ita\_Word
* **Variable(s) of interest are found:** Rating

Goals: we will use this data to provide examples of our simulation process on how to determine sample size for a project based on item rather than participant. You can read about this idea here: <https://github.com/SemanticPriming/SPAML/blob/master/02_Power/power_aipe.pdf> We will use the example provided in this link as the main portion of the paper and then add your data as a vignette example to supplement the paper. You will be considered an author for completing this template worksheet (no coding skills necessary, we will do that part), and reviewing/commenting on the draft of the paper. Please email [007spaml@gmail.com](mailto:007spaml@gmail.com) if you have questions.