Italian Age of Acquisition Norms for a Large Set of Words (ItAoA)

Mahmoud M.Elsherif

2023-09-11

### Vignette Setup:

### Project/Data Title:

Italian Age of Acquisition Norms for a Large Set of Words (ItAoA)

Data provided by: Ettore Ambrosini

### Project/Data Description:

The age of acquisition (AoA) represents the age at which a word is learned. This measure has been shown to affect performance in a wide 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 to 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 using adult estimates (Barca et al., 2002; Ferrand et al., 2008; Moors et al., 2013). Here, adult participants are asked to provide AoA ratings on a Likert scale (Schock et al., 2012; Alonso et al., 2015; Borelli et al., 2018) or directly in years, 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, the latter method is easier for participants to use and does not restrict artificially the response range, instead providing more precise information on the AoA of 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 (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 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:

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 consistency with these data collections (Montefinese et al., 2014; Fairfield et al., 2017). Word stimuli were distributed on 20 lists containing 97–98 words each. To avoid primacy or recency effects, the order in which words appeared on 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 someone 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:

Included with the vignette and Data Location: <https://osf.io/rzycf/>

DF <- import("ambrosini\_data.csv")  
  
DF <- DF %>%  
 arrange(Ita\_Word) %>% #orders the rows of the data by the target\_name column  
 group\_by(Ita\_Word) %>% #group by the target name  
 transform(items = as.numeric(factor(Ita\_Word)))%>% #transform target name into a item  
 select(items, Eng\_Word, Ita\_Word, everything()  
 ) #select all variables from items and target\_name   
  
DF <- DF %>%   
 group\_by(Ita\_Word) %>%  
 filter (Rating != 'Unknown')  
  
head(DF)

## # A tibble: 6 × 5  
## # Groups: Ita\_Word [1]  
## items Eng\_Word Ita\_Word SS\_ID Rating  
## <dbl> <chr> <chr> <int> <chr>   
## 1 1 dazzle abbaglio 282 16   
## 2 1 dazzle abbaglio 283 10   
## 3 1 dazzle abbaglio 284 12   
## 4 1 dazzle abbaglio 285 10   
## 5 1 dazzle abbaglio 286 8   
## 6 1 dazzle abbaglio 287 9

### Date Published:

2019-02-13

### Dataset Citation:

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>

### 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:

metadata <- import("ambrosini\_metadata.xlsx")  
  
flextable(metadata) %>% autofit()

| Variable Name | Variable Description | Type (numeric, character, logical, etc.) |
| --- | --- | --- |
| items | Item number | Numeric |
| Eng\_Word | English translation of the item | Character |
| Ita\_Word | Italian translation of the item | Character |
| SS\_ID | Subject ID Number | Numeric |
| Rating | Age of acquisition rating | Numeric |

### AIPE Analysis:

Note that the data are already in long format (each item has one row), and therefore, we do not need to restructure the data.

#### Stopping Rule

In this dataset, we have 48772 individual words to select from for our research study. You would obviously not use all of these in one study. Let’s say we wanted participants to rate 75 pairs of words during our study (note: this selection is completely arbitrary).

random\_items <- unique(DF$items)[sample(unique(DF$items), size = 75)]  
  
DF <- DF %>%   
 filter(items %in% random\_items)  
  
# Function for simulation  
var1 <- item\_power(data = DF, # name of data frame  
 dv\_col = "Rating", # name of DV column as a character  
 item\_col = "items", # number of items column as a character  
 sample\_start = 20,   
 sample\_stop = 100,   
 sample\_increase = 5,  
 decile = .4)

What is the usual standard error for the data that could be considered for our stopping rule using the 40% decile?

# individual SEs  
var1$SE

## 24 27 29 99 104 109 111 114   
## 0.6574699 0.4308132 0.4722993 0.4497407 0.5314132 0.5715476 0.4868949 0.4013311   
## 118 201 287 343 376 394 439 451   
## 0.4247352 0.6823489 0.2516611 0.3109126 0.6631239 0.3646002 0.5057008 0.5540156   
## 458 469 483 500 509 521 553 560   
## 0.2242023 0.4028234 0.3302524 0.5102940 0.5867424 0.6327717 0.4214262 0.6161169   
## 621 628 635 677 707 780 809 816   
## 0.2708013 0.4623130 0.4621688 0.3265986 0.2444040 0.5794250 0.6189238 0.4434712   
## 822 840 841 898 912 918 924 938   
## 0.4062840 0.3872983 0.5461380 0.3415650 0.4735680 0.3720215 0.5295281 0.6574699   
## 985 986 1008 1022 1034 1041 1062 1089   
## 0.2928026 0.2835489 0.5781580 0.6715157 0.4206344 0.4200000 0.4082483 0.3015515   
## 1096 1155 1187 1223 1227 1248 1256 1282   
## 0.2091252 0.3083288 0.2948446 0.6582806 0.3824483 0.6193545 0.2581989 0.2690725   
## 1341 1359 1395 1429 1435 1467 1469 1515   
## 0.3282276 0.5224302 0.5540156 0.5326662 0.5605949 0.3316625 0.4541143 0.4504812   
## 1522 1528 1538 1606 1617 1642 1650 1655   
## 0.4393935 0.4512206 0.3969887 0.5595236 0.3214550 0.4446722 0.5648599 0.2715388   
## 1660 1835 1865   
## 0.3390182 0.3229035 0.3555278

var1$cutoff

## 40%   
## 0.4048997

Using our 40% decile as a guide, we find that 0.405 is our target standard error for an accurately measured item.

#### Minimum Sample Size

To estimate the minimum sample size, we should figure out what number of participants it would take to achieve 80%, 85%, 90%, and 95% of the SEs for items below our critical score of 0.405?

cutoff <- calculate\_cutoff(population = DF,   
 grouping\_items = "items",  
 score = "Rating",  
 minimum = as.numeric(min(DF$Rating)),  
 maximum = as.numeric(max(DF$Rating)))  
# showing how this is the same as the person calculated version versus semanticprimeR's function  
cutoff$cutoff

## 40%   
## 0.4048997

final\_table <- calculate\_correction(  
 proportion\_summary = var1$final\_sample,  
 pilot\_sample\_size = DF %>% group\_by(items) %>% summarize(n = n()) %>%   
 pull(n) %>% mean() %>% round(),  
 proportion\_variability = cutoff$prop\_var  
 )  
  
flextable(final\_table) %>%   
 autofit()

| percent\_below | sample\_size | corrected\_sample\_size |
| --- | --- | --- |
| 85.33333 | 45 | 42.95703 |
| 85.33333 | 45 | 42.95703 |
| 90.66667 | 55 | 54.62725 |
| 98.66667 | 70 | 70.36361 |

Our minimum sample size is small at 80% (*n* = 43 as the minimum). We could consider using 90% (*n* = 55) or 95% 70.

#### Maximum Sample Size

While there are many considerations for maximum sample size (time, effort, resources), if we consider a higher value just for estimation sake, we could use *n* = 70 at 98%.

#### Final Sample Size

In any estimate of sample size, you should also consider the potential for missing data and/or unusable data due to any other exclusion criteria in your study (i.e., attention checks, speeding, getting the answer right, etc.). In this study, these values can be influenced by the other variables that we used to select the stimuli in the study.