# A Turkish Database for Psycholinguistic Studies Based on Frequency, Age of Acquisition, and Imageability

**Elif Ahsen Acar, Deniz Zeyrek, Murathan Kurfalı, Cem Bozşahin**
Cognitive Science Department
Informatics Institute, Middle East Technical University (METU), Ankara, Turkey
E-mail: ahsen.tolgay@metu.edu.tr, dezeyrek@metu.edu.tr, kurfali@metu.edu.tr, bozsahin@metu.edu.tr

**Abstract**

This study primarily aims to build a Turkish psycholinguistic database including three variables: word frequency, age of acquisition (AoA), and imageability, where AoA and imageability information are limited to nouns. We used a corpus-based approach to obtain information about the AoA variable. We built two corpora: a child literature corpus (CLC) including 535 books written for 3-12 years old children, and a corpus of transcribed children's speech (CSC) at ages 1;4-4;8. A comparison between the word frequencies of CLC and CSC gave positive correlation results, suggesting the usability of the CLC to extract AoA information. We assumed that frequent words of the CLC would correspond to early acquired words whereas frequent words of a corpus of adult language would correspond to late acquired words. To validate AoA results from our corpus-based approach, a rated AoA questionnaire was conducted on adults. Imageability values were collected via a different questionnaire conducted on adults. We conclude that it is possible to deduce AoA information for high frequency words with the corpus-based approach. The results about low frequency words were inconclusive, which is attributed to the fact that corpus-based AoA information is affected by the strong negative correlation between corpus frequency and rated AoA.

**Keywords**: psycholinguistic database, frequency, age of acquisition, imageability, corpus study

## 1 . Introduction

Psycholinguistic databases are reliable and practical sources for researchers since they provide standardized stimuli for scientific queries. They allow researchers to access a large pool of data and manipulate variables that suit their experimental purposes.

Linguistic variables are usually language-specific. For Turkish, resources for psycholinguistic variables exist but are scarce. In this study, we aim to contribute to the existing resources. We first describe the preliminaries of a psycholinguistic database containing ratings on age of acquisition (AoA) and imageability obtained from adult participants, and frequency, lexical category, and number of letters of words obtained from corpora.[1] Then, we offer a unique corpus-based method to obtain AoA values. A corpus-based method was first used by Carroll and White (1973) who took (English) word frequency counts from corpora of children's vocabulary (assembled by Rinsland, 1945; Dale, 1948; Dale & Eichholz, 1960) to assign AoA values to words that should be known in reading and used in writing by grade school children. The authors used these frequency counts for validity testing, which was affirmative. The corpora they used were designed for teachers to assist them in teaching vocabulary to children (i.e., the words in the corpora are words that are expected to be taught to children at a certain age, rather than the words that are reported to be known by them). The results of our study may be useful for researchers who deal with teaching Turkish to children as first or second language as well as for psycholinguistic researchers who need standardized stimuli.

In the rest of the paper; § 2 gives the background on the variables, § 3 explains the method devised for this study, § 4 presents and explains the results of the study, § 5 summarizes and concludes the paper.

## 2 . Background on the Variables

This section explains what word frequency, AoA and imageability refer to and discusses how they are typically obtained.

### 2.1 Word Frequency

There are two ways of obtaining word frequency in the related literature: objective and subjective frequency counts. Objective frequency is the actual number of times a word occurs in a specific text data or transcribed speech data. It can be extracted directly from the data through counting algorithms, hence the objectivity. Subjective frequency, on the other hand, refers to the estimates that are collected by asking participants to rate the words on the basis of their daily interactions with the words. How often the participants come across a word determines the subjective frequency rating of that word (Gernsbacher, 1984). In the current study, only objective frequency counts are used: the term *frequency* indicates objective frequency throughout the paper.

1 The database is freely available to researchers upon request via e-mail at *ahsen.tolgay@metu.edu.tr*

### 2.2 Age of Acquisition (AoA)

The AoA of a word indicates the age in which a child acquires that word. Acquisition does not necessarily require being able to use a word; comprehension in a context, or being able to define a word may also be different definitions of word acquisition (Ghyselinck et al., 2000). Research has found that high frequency words tend to be acquired earlier in life than low frequency words (Morrison et al., 1997; Ghyselinck et al., 2000; Bird et al., 2001), but there are exceptions. For instance *puppy* is a low frequency word that is acquired early in life, and *income* is a high frequency word that is acquired late (Morrison & Ellis, 1995).

The AoA variable is often divided into two: objective AoA and rated AoA. Objective AoA refers to the actual AoA obtained from child participants using the picture-naming paradigm, whereas rated AoA is obtained by asking adult participants to estimate their ages in the time of acquisition (Morrison et al., 1997). The reliability assessments of the rated AoA method have been conducted by using Ebel's method (Carroll & White, 1973), Cronbach's alpha (Rubin, 1980), or intergroup reliability (Gilhooly & Hay, 1977; Gilhooly & Logie, 1980a; Winters et al., 1978), all of which have pointed to a reliable rating. Several validity studies have also been conducted (e.g. Gilhooly & Gilhooly, 1980; Morrison et al., 1997). Rated AoA has thus been accepted as a reliable and valid measure for the AoA variable.

### 2.3 Imageability

Imageability refers to "the ease with which a word arouses sensory images" (Paivio et al., 1968, pp. 2). For instance, when a person encounters the word *blanket*, an image of a blanket forms easily and instantly in the mind. This indicates that *blanket* is a highly imageable word and would have a higher rating on an imageability scale. On the other hand, the word *honour* does not form a mental image as easy as *blanket*, meaning that *honour* would have a lower value on the scale. Typically, imageability ratings are obtained by asking participants to fill a questionnaire by rating a number of words on a scale (Paivio et al., 1968). There is a difference between the terms imageability and concreteness, which refers to the actual sensory experience of the objects. It has been found that words such as *anger*, which have low values on concreteness scales might have high imageability (Paivio et al., 1968). Nevertheless, some researchers prefer to use the terms imageability and concreteness interchangeably because of the strong correlation found between them (Paivio et al., 1968, Reilly & Kean, 2007). In the current study, we use the concept imageability rather than concreteness.

Research has further shown that imageability and AoA are intercorrelated (e.g. Coltheart et al., 1988; Cortese & Fugett, 2004; Stadthagen-Gonzalez & Davis, 2006; Ma et al., 2009). For example for Chinese, it was found that noun and verb AoA could be predicted by imageability (Ma et al., 2009). The relationship between imageability and AoA was also found in child reading tasks (Coltheart et al., 1988). These studies point out a close relationship of imageability and AoA, confirming the need to consider imageability together with frequency and AoA.

## 3 . Method

This section describes the method used in the study, namely, the corpora we built or collated, the questionnaires we developed, and the parsing and part-of-speech (POS) tagging of the corpora we carried out.

### 3.1 The Corpora

We built two kinds of corpora on the basis of which we determined word frequency and AoA: (a) a corpus of child literature (CLC) comprising books written for Turkish children between ages 3-12, and (b) a corpus of children's speech (CSC). The CLC is formed on the assumption that it is possible to deduce AoA information from children's books. It contains 535 children's books written by Turkish authors amounting to 4,388,149 word tokens, 19,246 word types. The CSC includes children's speech from the Turkish sections of the CHILDES database (MacWhinney, 2000).[2] In addition, it incorporates children's speech collected from Kindergarten students (1;4-4;8 years of age) in Ankara, amounting to a total of 33,854 word tokens and 1,912 word types. In addition to these, a web-based corpus, namely, a part of the BOUN Corpus (Sak et al., 2011), was used as a corpus representing adult language. This is a compilation of written Turkish articles from online news portals and general web pages containing approximately 490 million tokens. We used a random section of this corpus, which we refer to as the BOUN sub-corpus, consisting of 2,832,025 word tokens and 11,349 word types.

Our method involves a comparison of word frequencies in the CLC and the BOUN sub-corpus to determine early and late word AoA information and checking the validity of this information with data collected from adult participants. We first compared word frequencies in books for 3-5 year-old children with those in the CSC. There were 1,060 words that were listed in both the CSC and this section of the CLC; the comparison analysis was made using these words' frequency counts. We hypothesized that a high correlation would indicate the usability of the CLC section as a representative of child speech, although the higher age ranges of CLC could not be compared to any speech data due to the scarcity of available resources on Turkish child speech. Secondly, we compared CLC word frequencies to the BOUN sub-corpus frequencies. We predicted that this comparison would provide a distinction between early and late acquired words. The words that exist in both corpora were extracted and divided into two categories based on a

2 We used the CHILDES files created by Ayhan Aksu-Koç in 2004 and Feyza Turkay in 2012.

comparison between each word's own frequency count in its respective corpus: the words whose CLC frequency counts are higher than BOUN sub-corpus frequency counts (referred to as the CLC>BOUN words) were assumed to be early acquired, thus expected to have low AoA ratings from the questionnaire data, whereas the words whose BOUN sub-corpus frequency counts are higher than CLC frequency counts (referred to as the BOUN>CLC words) were assumed to be late acquired, and expected to have high AoA ratings from the questionnaire data. The amount of words in these categories was 8,844 in total.

We tested our corpus-based assumptions on AoA by comparing them to the AoA ratings collected from adults. For this comparison, we only dealt with nouns. The questionnaire we developed to determine adults' AoA ratings and imageability levels are explained in the next section.

## 3.2 AoA and Imageability Questionnaires

In order to collect rated AoA, a questionnaire was created using 300 nouns from the BOUN sub-corpus as well as the CLC. The same set of words was used to create the imageability questionnaire. Both questionnaires were distributed to participants via a website designed for surveys (http://www.qualtrics.com/).

### 3.2.1 Materials for the Questionnaires

All the words were selected among uninflected common nouns in the corpora. Compound words were manually excluded from the selection because they are morphologically complex words, which are processed differently than single words (e.g. Fiorentino & Poeppel, 2007). 256 words were randomly chosen from the CLC and the BOUN sub-corpus keeping balanced samples from high and low frequency words. The frequency counts per million words were first transformed with the base-10 logarithm function in order to achieve a normal distribution; then they were sectioned using quartile analysis, where the same number of words was randomly chosen from each quartile.[3] Of these 256 words, 192 came from the words that existed in both corpora (half of them is from CLC>BOUN words, while the other half is from BOUN>CLC words), 32 came from the words that existed in CLC but not in the BOUN sub-corpus, and 32 came from the words that existed in the BOUN sub-corpus but not in CLC. The purpose was to keep the number of items reasonable for a questionnaire and have a balanced frequency sample from each corpus. In addition, 44 control words were randomly selected from a set of 260 words containing AoA ratings collected by Raman et al. (2014). The control words were used in order to check the participants' consistency of response, i.e., to see whether they gave random ratings to the questionnaire items or not (Kuperman et al., 2012).

### 3.2.2 Participants

Forty-seven adult native speakers of Turkish participated in the AoA questionnaire. Four of the participants' data were not included in the analysis because they included inconsistent AoA information, i.e. the AoA they entered was higher than their age in the demographic forms. Of the 43 remaining participants, 28 were female and 15 were male. Their age range was 19 to 55; the mean age was 28.63. The education levels were as follows: 4 had an associate degree, 1 had a bachelor's degree, 25 had a master's degree, and 13 had a doctorate degree.

Twenty-eight adult native speakers of Turkish participated in the imageability questionnaire, with 17 females, 11 males. The age range was 18 to 48, and the mean age was 28.43. The education levels were as follows: 1 was high school graduate, 21 had a bachelor's degree, and 6 had a master's degree.

All the participants took part in the questionnaires voluntarily.

### 3.2.3 Procedure

The questionnaires were conducted online. They were made available to participants via social media and e-mail groups. Participants were required to read a consent form at the beginning of each questionnaire and accept the terms. Then they had to read the explanations on how to proceed.

In the AoA questionnaire, first the term AoA was described. Then, 8 sample words and their AoA ranges from Raman et al. (2014) were presented. The sample words were different from the control words. These can be referred to as calibrator words because they help the participants calibrate their responses (Kuperman et al., 2012). The participants were instructed to enter the age they think they had learned the words, and '0' for any words they did not know. They were required to enter the numbers manually in a text box below each item. After these instructions, the word lists were presented in 6 pages with 50 words in each (300 words in total). The word order was randomized by the Qualtrics website for each participant. At the end of the questionnaire there was a demographic information form.

In the imageability questionnaire, the participants were presented with instructions partially adapted from Paivio et al. (1968). First, the term imageability was described with

3 Since the comparison analyses were conducted using frequencies obtained from two distinct corpora, we needed a standardized measure to balance the difference between the sizes of each corpus. The frequency per million words (fpmw) measure did not yield a normal distribution of frequency counts; the distribution was highly skewed. Therefore, a log10 transformation was applied to the fpmw values which is generally regarded to be adequate to standardize the frequency counts. Van Heuven et al. (2014) propose a new scale, called the Zipf scale. It is basically presented as a 7 point scale in which low-frequency values are 1 (0.01 fpmw), 2 (0.1 fpmw) and 3 (1 fpmw), whereas high-frequency values are 4 (10 fpmw), 5 (100 fpmw), 6 (1,000 fpmw), and 7 (10,000 fpmw). The authors report that the Zipf scale eliminates the negative values that result from the logarithm transformation.

two examples. Then, the participants were instructed to assign a number between 1 (for easiest imageability) and 7 (for hardest or no imageability) to rate the imageability of words. They were asked to rate only the presented words and not any associated words. After the instructions, the word lists were presented in 6 pages with 50 words in each. The word order was randomized by Qualtrics for each participant. At the end of the questionnaire, the same demographic information form used for the AoA questionnaire was presented.

### 3.3 Parsing the Corpora

Turkish is an agglutinating language with derivational and inflectional suffixes added to the word root. We morphologically parsed the CLC and the BOUN sub-corpus using the parser and disambiguator of Sak et al. (2007, 2008). Our purpose was two-fold: to extract uninflected nouns for use in our questionnaires, and to include the outputs of the parser (word frequency, lexical categories) in a database that also includes AoA and imageability ratings from or work. The morphologically parsed data is an output of our work that can also be used in future Turkish NLP studies.

The parser provides various possible part-of-speech (POS) tags for a given word and the disambiguator chooses the correct parse from the possible options. The parsing performance should have an accuracy of 87.67 % according to Eryiğit (2012); our parsing analyses on the CLC yielded an accuracy of 69%. To improve the accuracy rate, we extracted 5,026 words from 26 books in the CLC, selected on a balanced distribution of the age groups and the number of words each book contains, then manually corrected the POS tags of these words. We used them in a new training data set for the disambiguator. For some words it was not possible to assign the grammatically correct POS tags due to compatibility issues with the disambiguator. Therefore, our morphologically parsed CLC section (5,026 words) can be referred to as a silver standard data. The new training data containing the silver standard data increased the accuracy to 87.93% on the CLC. The BOUN sub-corpus was parsed and disambiguated without this additional manual work on the training data.

## 4 . Results

### 4.1 Word Frequencies in CLC and CSC

Figure 1 shows the relationship of log 10 transformed frequency values between child speech (ages 1;4-4;8) and a comparable part of the CLC (i.e., books targeting children at ages 3-5). There is a statistically significant positive correlation between them ($r(1060) = 0.58$, $p<0.01$), supporting the hypothesis that the CLC section under investigation is representative of children's speech.

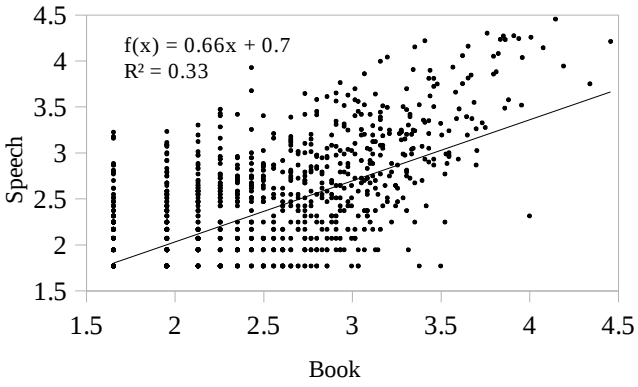


Figure 1. The relationship between frequency values in child speech and children's books

### 4.2 Corpus-Based AoA and Rated AoA Values

The correlation analysis for comparing the control words and Raman norms (Raman et al., 2014) revealed a strong positive correlation ($r(44) = 0.84$, $p<0.01$). This result points to the reliability of the participant responses. Moreover, split-half reliability was checked. The questionnaire was found to be highly reliable (22 items, Cronbach's $\alpha = 0.97$).

| | *CLC frequency> BOUN frequency (Early acquired)* | *BOUN frequency> CLC frequency (Late acquired)* |
|---|---|---|
| *Above 50%* | -0.39 (0.006) | -0.44 (0.002) |
| *Below 50%* | -0.23 (0.067) | -0.32 (0.029) |

Table 1. Pearson's correlation coefficients between AoA ratings and corpus-based AoA for high- and low-frequency nouns for shared items in the CLC and the BOUN sub-corpus (p-values are reported in parentheses).

In order to test the assumptions that nouns from CLC>BOUN words would have lower AoA ratings and nouns from BOUN>CLC words would have higher AoA ratings, a correlation analysis was performed for each quartile. This did not reveal any significant results, probably because the number of nouns in each quartile (24) was low. Therefore, another correlation analysis was performed for 50% segments in both lists, assuming that this would give a distinction between high and low frequency nouns. The result of this analysis is summarized in Table 1. The analysis shows that the frequency counts of nouns above the 50% line (above 50% nouns) from the CLC>BOUN words were negatively correlated with their respective AoA ratings, which means that the participants estimated the high frequency nouns to be acquired early in life ($r(48) = -0.39$, $p<0.01$). This is consistent with our hypothesis. However, the nouns above the 50% line from

the BOUN>CLC words did not get high AoA ratings, as indicated by the negative correlation between them ($r(48) = -0.44$, $p<0.01$). They were also estimated by the participants to be acquired early in life, contrary to our assumption.

The nouns below the 50% line (below 50% nouns) from CLC>BOUN words did not have a significant correlation result, although there was a negative correlation between frequency counts and AoA ratings ($r(48) = -0.39$, $p>0.05$). On the other hand, the negative correlation between the frequency counts of below 50% nouns from the BOUN>CLC words and AoA ratings was significant ($r(48) = -0.32$, $p<0.05$). This suggests that these nouns were acquired later in life as expected.

### 4.3 Frequency and Rated AoA

For a more general comparison between corpus word frequencies and rated AoA, a correlation analysis without frequency distinctions was conducted. There was a strong negative correlation between rated AoA and CLC frequencies ($r(224) = -0.70$, $p<0.01$), while the correlation between rated AoA and the BOUN sub-corpus frequencies was not as strong ($r(224) = -0.29$, $p<0.01$). These findings are compatible with research which has shown that high-frequency words have low AoA values (e.g., Ghyselinck et al., 2000). Multiple regression analyses also supported the results of the correlation analyses, and the direction and significance of these relationships did not change. These results show that nouns reported to be acquired early in life are usually high-frequency nouns no matter which source they were extracted in our study (i.e. nouns above the 50% line in both corpora). Statistically, the relationship between frequency and AoA is found to be strong. It may be due to this reason that the AoA distinction we expected to observe between the nouns in the BOUN sub-corpus and the CLC were obscured. In further work, the use of the Zipf scale (van Heuven et al., 2014) could resolve part of the issues that arose from the negative correlation between frequency and AoA variables in our study. For instance, the nouns in the high- and low-frequency categories could be analysed separately from the nouns between scales 3 and 4.

### 4.4 Frequency, Rated AoA, and Imageability

We checked split-half reliability in the imageability questionnaire. The questionnaire was found to be highly reliable (14 items, Cronbach's $\alpha = 0.92$).

We explored the relationships between imageability, rated AoA, and CLC and BOUN sub-corpus frequencies. All these calculations are based on nouns. We found that imageability ratings are highly correlated with rated AoA values ($r(300) = -0.77$, $p<0.01$). As imageability increased, rated AoA decreased. This is in agreement with previous research (e.g., Stadthagen-Gonzalez & Davis, 2006) indicating that highly imageable words tend to have lower AoA ratings, suggesting an early acquisition. High imageability, then, points to low rated AoA values just as high frequency does. Furthermore, we found that as imageability increases, the CLC word frequencies also increase ($r(224) = 0.41$, $p<0.01$). However, the correlation between imageability and the BOUN sub-corpus frequency counts was low with regard to that between imageability and CLC frequencies ($r(224) = 0.19$, $p<0.01$). Further analyses with multiple regression revealed that when the effect of rated AoA was suppressed, the significant relationship between the BOUN sub-corpus frequency and imageability disappeared. Rated AoA significantly predicted imageability ($\beta = -0.70$, $t(218) = -13.70$, $p<0.01$), but the BOUN sub-corpus frequencies did not significantly predict imageability ($\beta = -0.02$, $t(218) = -0.36$, $p = 0.73$). Moreover, while the CLC frequency and imageability relationship remained significant, the direction of the relationship changed. The analysis showed that rated AoA ($\beta = -0.76$, $t(218) = -10.97$, $p<0.01$) and the CLC word frequency ($\beta = -0.14$, $t(218) = -2.04$, $p < 0.05$) significantly predicted imageability. Without the effect of rated AoA, as the CLC word frequency increased, imageability decreased. In other words, without the effect of rated AoA, the low-frequency nouns from the CLC have higher imageability ratings than the high-frequency nouns, which is contrary to the correlation results reported at the beginning of this sub-section. The strong correlation between rated AoA and imageability ($r(300) = -0.77$, $p<0.01$) might be a reason for this change. There was also a strong correlation between CLC frequency and rated AoA ($r(224) = -0.70$, $p<0.01$), which is another probable effect on this change.

## 5 . Summary and Conclusions

We created the preliminaries of a psycholinguistic database for Turkish that includes frequency values per million words as well as information on lexical category and number of letters. The database is based on 19,246 word types from the CLC, and 11,349 word types from the BOUN sub-corpus, with frequency values of a total of 8,844 common word types obtained from two corpora.[4] The database also includes rated AoA and imageability values for 300 nouns obtained by means of the questionnaires we developed.

We used a corpus-based method to obtain frequency values and deduce AoA information from written corpora. The ideal way for this would be to use spoken language corpora of adults and children but this was not possible due to scarcity of available resources on Turkish. Therefore, the CLC was created to represent children's language. The BOUN sub-corpus was used as a representative of the adult language. A correlation analysis of the frequencies between children's speech (our CSC) and the CLC was performed,

4 The full content of the CLC cannot be shared under any circumstances due to copy-right contracts between the authors and the publishers.

yielding a significantly positive result, supporting the assumption that the corpus of children's books we formed was a reliable sample of children's language.

We assumed that high-frequency words from CLC would be early acquired words, thus have low AoA values; whereas high-frequency words from the BOUN sub-corpus would be late acquired, thus have high AoA. The results of the analyses showed us that the above 50% nouns from CLC>BOUN words got low AoA ratings in the questionnaire. This was as expected. However the above 50% nouns from BOUN>CLC words did not get high AoA ratings contrary to expectations. We attribute this result to the negative correlation between frequency and rated AoA, which acted as a confounding factor. On the other hand, below 50% nouns from both CLC>BOUN and BOUN>CLC words displayed the expected tendencies with respect to AoA (as frequency increases, AoA decreases). However, the correlations of CLC>BOUN nouns with rated AoA were not significant. The problem may be due to the fact that most participants could not rate the low-frequency nouns (in terms of AoA) because they simply did not know them. In the future, increasing the number of participants will be necessary to ensure adequately many participants who rate low-frequency items informatively, i.e. with values other than a simple '0'.

To sum up, it was possible to infer AoA information from above 50% nouns in CLC>BOUN, largely confirming the validity of our corpus based approach. These nouns were found to be early acquired. Our attempt to infer AoA information on the basis of above 50% nouns from BOUN>CLC did not give expected results. Regarding imageability, we found that imageability ratings are highly correlated with rated AoA values; i.e. as imageability increased, rated AoA decreased. The CLC frequency had a strong positive correlation with imageability, while the BOUN sub-corpus correlation with imageability was not as strong. Moreover, multiple regression analyses revealed that rated AoA had a significant effect on the relationship between imageability and CLC frequency, and also between imageability and BOUN sub-corpus frequency.

## 6 . Acknowledgements

The following publishers deserve special gratitude and appreciation for their indispensable contributions to the CLC: TUDEM Yayın Grubu, YKY Doğan Kardeş Yayınları, Mavibulut Yayınları, Pencere-sey, 3P Yayıncılık. Moreover, we appreciate Gökhan Gönül's assistance in collecting and transcribing the data for the CSC.

## 7 . Bibliographical References

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers 33:1*, 73-79.

Carroll, J. B., & White, M. N. (1973). Word frequency and age-of-acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology 25:1*, 85-95.

Coltheart, V., Laxon, V. J, & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology 79*, 1-12.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers 36:3*, 384-387.

Dale, E. (1948). Dale list of 3,000 familiar words. Educational Research Bulletin 27,45-54.

Dale, E. & Eichholz, G. (1960). *Children's Knowledge of Words.* Columbus, Ohio: Bureau of Educational Research and Service, Ohio State University.

Eryiğit, G. (2012). *The Impact of Automatic Morphological Analysis & Disambiguation on Dependency Parsing of Turkish.* In proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 12). Istanbul, Turkey.

Fiorentino, R. & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes 22:7*, 953-1000.

Gernsbacher, M. A. (1984). Resolving 20 Years of Inconsistent Interactions Between Lexical Familiarity and Orthography, Concreteness, and Polysemy. *Journal of Experimental Psychology: General 113:2*, 256-281.

Ghyselinck, M., De Moor, W., & Brysbaert, M. (2000). Age-of-acquisition ratings for 2816 Dutch four- and five-letter nouns. *Psychologia Belgica 40:2*, 77-98.

Gilhooly, K. J., & Gilhooly, M. L. M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology 71*, 105-110.

Gilhooly, K. J., & Hay, D. (1977). Imagery, concreteness, age-of-acquisition, familiarity, and meaningfulness values for 205 five-letter words having single-solution anagrams. *Behaviour Research Methods & Instrumentation 9*, 12-17.

Gilhooly, K. J., & Logie, R. H. (1980a). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behaviour Research Methods & Instrumentation 12:4*, 395-427.

Ma, W., Golinkoff, R. M., Hirsh-Pasek, K., McDonough, C. & Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of Child Language 36:2*, 405-423.

Kuperman, V., Stadthagen-Gonzalez, H. & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods 44:4*, 978-990.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory & Cognition, 21:1*, 116-133.

Morrison, C. M., Chappell, T. D. & Ellis, A. W. (1997). Age of Acquisition Norms for a Large Set of Object Names and Their Relation to Adult Estimates and Other Variables. *The Quarterly Journal of Experimental Psychology 50:3*, 528-559.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 words. *Journal of Experimental Psychology Monograph Supplement, 76(3, Pt. 2)*.

Raman, I., Raman, E., & Mertan, B. (2014). A standardized set of 260 pictures for Turkish: Norms of name and image agreement, age of acquisition, visual complexity, and conceptual familiarity. *Behavior Research Methods 46*, 588-595.

Reilly, J. & Kean, J. (2007). Formal Distinctiveness of High- and Low-Imageability Nouns: Analyses and Theoretical Implications. *Cognitive Science 31*, 157-168.

Rinsland, H. D. (1945). *A Basic Vocabulary of Elementary School Children*. New York: Macmillan.

Rubin, D.C. (1980). 51 properties of 125 words: A unit analysis of verbal behaviour. *Journal of Verbal Learning and Verbal Behaviour 19*, 736-755.

Stadthagen-Gonzalez, H. & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods 38:4*, 598-605.

van Heuven, W. J. B., Mandera, P., Keuleers, E. & Brysbaert M. (2014) SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology 67:6*, 1176-1190.

Winters, J.J., Winter, L., & Burger, A.L. (1978). Confidence in age-of-acquisition estimates and its relationship to children's labelling performance. *Bulletin of the Psychonomic Society 12*, 361-364.

## 8 . Language Resource References

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Sak, H., Güngör, T., & Saraçlar, M. (2007). *Morphological disambiguation of Turkish text with perceptron algorithm*. In CICLing 2007, volume LNCS 4394, pages 107-118.

Sak, H., Güngör, T., & Saraçlar, M. (2008). *Turkish Language Resources: Morphological Parser, Morphological Disambiguator and Web Corpus*. In GoTAL 2008, volume 5221 of LNCS, 2008, pages 417–427. Springer.

Sak, H., Güngör, T., & Saraçlar, M. (2011). Resources for Turkish morphological processing. *Language Resources & Evaluation 45:2*, 249-261.