

# Demographic influences on word associations

Marta Marija Aleknavičiūtė

Vilnius University

[marta.aleknaviciute@gmail.com](mailto:marta.aleknaviciute@gmail.com)

ORCID: <https://orcid.org/0009-0008-8136-5074>

**Abstract.** Word associations are verbal responses to cues, typically studied using word association tests. This study examines how demographic factors, such as gender, age, settlement size, ethnographic region and education level, influence word associations. Data were collected online using a 30-stimulus research tool, resulting in 11,046 responses from 389 participants. Stereotypy, defined as the tendency for the same response to recur for a given cue word, was analyzed using multiple regression. The findings indicate that the associations provided by younger participants were more stereotypical than those of older participants, and women's responses exhibited a higher degree of stereotypy compared to men's. However, other demographic factors did not show a significant impact. The findings suggest that demographic factors help shape word associations, contributing to the growing body of research that highlights the influence of various factors on associative patterns. Additionally, this study provides valuable insights into Lithuanian, an under-researched language in the field of word association studies.

**Key words:** *word associations, word association test, demographic factors, multiple regression*

## Demografinių veiksnių įtaka žodžių asociacijoms

**Santrauka.** Žodžių asociacijos – tai ryšys tarp stimulo ir reakcijos, dažniausiai išgaunamos žodžių asociacijų testu. Žodžių asociacijų testas yra eksperimentinis tyrimo metodas, vykdomas taip: dalyviams pateikiamas žodis–stimulas ir prašoma atsakyti pirmu mintyse kilusiu žodžiu. Dažnai žodžių asociacijų tyrimų rezultatas būna vadinamieji asociacijų normų sąrašai – tai sąrašai, žodynai ar duomenų bazės, kuriose nurodoma, kaip dažnai tam tikri stimulai sukelia vienokias ar kitokias žodžių asociacijas, kurie esą yra tipiški visai kalbinei bendruomenei. Kita vertus, pastaruoju metu pradėtas kelti klausimas, ar tokių normų sąrašų sudarymas yra tikslingas, nes žodžių asociacijos varijuoja. Manoma, kad žodžių asociacijoms įtaką daro stimulus ir eksperimento dalyvis. Šiame tyrime nagrinėjama dalyvių įtaka – tiriama demografinių veiksnių įtaka žodžių asociacijoms. Siekiama atsakyti į klausimą, ar demografiniai veiksniai, tokie kaip lytis, amžius, gyvenamosios vietos dydis, etnografinis regionas ir išsilavinimas, daro įtaką asociacijoms ir, jei daro, kokią. Tikslui pasiekti sudarytas tyrimo įrankis, kurį sudaro 30 abstrakčių stimulų ir sociolingvistinis klausimynas. Eksperimente dalyvavo 389 dalyviai, kurie pateikė 11 046 asociacijas, kurios sudarė tiriamąją medžiagą. Demografinių veiksnių įtaka asociacijoms buvo tiriama kiekybiniu metodu – atlikta regresinė analizė parodė, kad skirtingų demografinių charakteristikų dalyviai pateikė skirtingas asociacijas: jaunesni tyrimo dalyviai dažniau pateikė stereotipiškas asociacijas negu vyresni tyrimo dalyviai, o moterys tokių asociacijų pateikė daugiau negu vyrai. Nustatyta, kad kiti tirtieji demografiniai veiksniai tokios įtakos nedarė. Tokie tyrimo rezultatai rodo, kad žodžių asociacijų normų sąrašai galimai neatspindi kalbinės bendruomenės heterogeniškumo ir tyrimuose vertėtų atsižvelgti į demografinių veiksnių įtaką žodžių asociacijoms.

**Raktiniai žodžiai:** *žodžių asociacijos, žodžių asociacijų testas, demografiniai veiksniai, daugianarė regresija*

## 1. Introduction

### 1.1. Studies on word associations

The concept of word associations is often explained through the description of the word association test, an experimental research method typically conducted as follows: participants are presented with a cue word (in written or spoken form) and are asked to respond with the first word that comes to mind. The result of such a test is a set of collected word associations, which can be used in various fields.

Word associations are studied by both psychologists (Jung 1910; Bandera et al. 1991; Petchkovsky et al. 2013) and linguists across disciplines. The latter examine word associations in various contexts, including foreign language teaching (Higginbotham 2010; Cremer et al. 2011; Fitzpatrick and Thwaites 2020), the mental lexicon (De Deyne and Storms 2008; De Deyne et al. 2019), corpus linguistics (Lin et al. 2019) and the priming effect (Zeelenberg et al. 2003). Overall, linguists generally agree that word association studies provide insights into lexical availability, organization and retrieval, contributing to a deeper understanding of the human mental lexicon.

The results of word association studies often lead to the creation of association norm lists (Nelson et al. 2004; Karimkhani et al. 2021), which may take the form of lists, dictionaries or databases indicating how frequently certain cues evoke specific word associations that are considered typical for a linguistic community. The lists are used, for example, to compare the word associations of individuals with dementia and healthy individuals (Bandera et al. 1991), or to analyze the differences between L2 users' associations and those typical for L1 users (Fitzpatrick 2009). Thus, they are useful for comparing certain individuals or groups against a reference group. However, some studies have questioned the validity of such norm lists (Fitzpatrick 2007). Since word association responses vary widely among participants, some researchers argue that such lists may not fully capture the heterogeneity of a linguistic community. The present study also attempts to delve into the question of how appropriate it is to rely on norm lists that generalize a linguistic community.

Since this study focuses on Lithuanian word associations, it is important to acknowledge the existence of two Lithuanian word association lists. Steponavičienė (1986) compiled a dictionary of word associations, while Vilkaitė-Lozdienė (2019a) initiated the development of a Lithuanian word association database. Unlike the former, Vilkaitė-Lozdienė's database is not in print and is freely available online, allowing other researchers to both use it and contribute their own data. The word associations collected in the present study will be added to this database, making them publicly available for future research.

Although a few other studies have examined word associations in Lithuanian (Papurėlytė 2011, 2012, 2014; Akelaitienė 2007), their primary focus has not been on word associations as a means of investigating the mental lexicon. Instead, these studies analyze the associative fields of specific lexical items – such as *medis* ('tree'), *žodis* ('word'), and *kaimas* ('village') – to explore how associations reflect the linguistic worldview. While gender differences were considered in some of these studies, the analyses were limited to qualitative semantic interpretation. In contrast, the present study adopts a quantitative approach, employing regression analysis to examine the influence of multiple demographic factors – including, but not limited to, gender – on word association patterns. Therefore, the **aim** of this study is to examine whether different demographic factors influence word associations within a single regression model, allowing for the contextual assessment of each factor's impact and, if so, to determine how these factors affect associative patterns. Prior to addressing this objective, it is necessary to review existing studies that examine the influence of various factors on word associations.

According to Fitzpatrick (2007: 323), word associations are influenced by two factors: the cue word and the participant. The significance of these factors may vary in different situations, but they always have some degree of influence on the associations. The following sections will examine these two factors in detail, beginning with the influence of cue words on word associations.

## 1.2. The influence of cue words on word associations

The impact of the cue on associations can vary. Aitchison (2003, cited in Schmitt 2010: 61) observed that associations almost always belong to the same semantic field as the cue. The part of speech of a cue word is also important. For example, Nissen and Henriksen (2006) found that nouns elicit more paradigmatic associations than adjectives and verbs, while adjectives and verbs elicit more syntagmatic associations than nouns<sup>1</sup>.

Beyond part of speech, morphology also affects word associations. Vilkaitė-Lozdienė (2019b) studied word associations for different morphological forms of verbs and nouns and found that the morphological form of Lithuanian verbs does not significantly affect word associations, whereas noun morphology does: nouns in the genitive and accusative cases elicit more syntagmatic associations than those in the nominative case.

Another important factor is word frequency. Meara (1983) found that high-frequency words elicit more predictable associations, which may not be ideal when studying individual differences. Similarly, Wolter (2001) observed that both L1 and L2 language users tend to provide more form-based or inconsistent associations when responding to lower-frequency words. Additionally, Vitrano et al. (2021) found that participants generated more associations for high-frequency words than for low-frequency ones. These findings highlight the importance of considering word frequency when conducting word association research.

Finally, the concreteness of a cue word also plays a crucial role in shaping word associations. De Groot (1989) found that abstract cues elicit a wider range of responses than concrete ones, possibly due to broader cognitive mechanisms shaping word meaning. Research suggests that abstract word meanings are grounded not only in linguistic input but also in sensorimotor experience, emotional experience and social interactions (Borghi et al. 2017). These factors may help explain why demographic groups exhibit distinct word association patterns, as diverse experiences shape their mental lexicon in unique ways, reinforcing the role of participants in word association studies.

These findings suggest that the cue word significantly impacts word associations, and it is crucial to choose cues that best serve the research objectives when studying word associations. Since this study aims to examine the influence of participants on word associations, it was assumed that the diversity of cue words in terms of linguistic characteristics is not as crucial; rather, using cues with similar characteristics would help minimize the cue influence on associations and focus on the impact of demographic factors. Therefore, only noun cues were used, presented to the participants in the nominative case. Additionally, abstract cue words were chosen to elicit a diverse range of associations.

---

<sup>1</sup> Paradigmatic associations are associations where the part of speech matches that of the cue, and they are linked by meaning or semantic closeness, for example, *black – white*. Syntagmatic associations are those that form a word combination with the cue and often belong to different parts of speech, for example, *black – dog*.

### 1.3. The influence of participants on word associations

While cue-word characteristics shape word associations, participant-related factors may also have an impact. Controlling for linguistic variation in cue words allows this study to focus more directly on demographic influences.

The most widely studied demographic factor affecting word associations is age. Studies show that word associations differ among people of different ages. It is reasonable to assume that children's word associations often differ from those of adults (Mattheoudakis 2011). However, associations provided by adults from different age groups may also differ. For example, Hirsh and Tree (2001) found that the younger age group (21–30 years) was more heterogeneous in terms of associations compared to the older group (66–81 years). The associations of younger participants were more diverse: fewer people gave the most common word associations, and there were more unique responses. Another study highlighting the differences in word associations across different age groups was conducted by Fitzpatrick et al. (2015). They studied twin pairs in two age groups: 16-year-olds and 65+ year-olds. The results showed no difference between twins of the same age, regardless of birth order. However, there was a significant difference by age – the participants' results were more similar to those of other participants in the same age group than to those in the other age group.

As Dubossarsky et al. (2017: 1) state, throughout life, a person is surrounded by language, which enriches the mental lexicon with new associative information over time. Their study showed that word associations become more predictable from childhood to age 30 but grow more diverse and less predictable afterward. Given its large sample and broad age range, this research provides strong evidence of the relationship between association patterns and age.

There is much less data on how the gender of the participants influences word associations. Aside from some studies conducted in the last century (Tresselt et al. 1955; Jenkins and Palermo 1965), few studies have addressed this issue. However, a recent study by De Deyne et al. (2023) identified gender-related differences in word meaning. Using a constrained word association task, participants were asked to provide three adjectives for each cue word. The findings suggest that the conceptual representation of common words varies between genders, highlighting the need for further research on the influence of various factors on word meaning and, consequently, word associations.

There is also very little data on how participants' education affects word associations. Many word association studies involve university students (Nelson et al. 2004; Playfoot et al. 2018), and such homogeneous groups in terms of education limit the ability to determine whether education affects word associations. This raises the question of whether education-related differences exist in word associations at all – a question that remains largely unanswered.

The current study aims to contribute to this knowledge by examining the influence of demographic factors such as gender, age, education and place of residence on word associations. The latter factor, to the author's knowledge, has not yet been studied. However, the present study suggests that this factor should not be overlooked. Given the extensive dialectal variation and widespread use of languages such as English, it is plausible that word associations may differ across speakers from different regions. If this study finds that place of residence significantly influences word associations in Lithuanian, it could suggest that similar patterns may emerge in larger and more widely spoken languages.

## 2. Methodology

### 2.1. Research instrument

To examine whether specific demographic characteristics influence word associations, participants first completed a sociolinguistic questionnaire before taking the word association test. Participants were asked to indicate their gender, age, place of residence (categorized by city size<sup>2</sup> and region) and educational attainment. All participants were required to give consent prior to participation. To ensure privacy and confidentiality, the data were collected anonymously. After completing the questionnaire, participants proceeded to the word association test.

The cue words were selected from the list of word frequencies in the Lithuanian language compiled by Dadurkevičius (2020), available in the CLARIN-LT open-access resource repository<sup>3</sup>. It was defined that abstract nouns would be selected from the first to the third thousand most frequent nouns in the list. The list of cue words can be found in the appendix.

Since the aim was to explore the participants' influence on associations, the diversity of cue words was not as crucial. For this reason, it was decided to study a relatively homogeneous group of cues. As discussed in the introduction, several factors influenced the selection of cues in this study. First, only nouns were selected for the study, and they were presented to the participants in the nominative singular form. Moreover, it was defined that only abstract nouns would be studied.

Participants were presented with 30 cue words, which is fewer than in some studies using around 100 cue words (Wettler et al. 2005; Playfoot et al. 2018). This decision was made because the data were collected online with voluntary participation, increasing the likelihood that some participants might not complete a longer experiment.

The data for this study were collected online using the cognitive-psychological experiment platform PsyToolkit (Stoet 2010, 2017). Participants were invited to join the study online via Facebook and were asked to provide the first word that came to mind after reading a cue word. To minimize the risk of priming effects, the order of cue words was randomized for each participant. This ensured that associations were not systematically shaped by the sequence in which stimuli appeared, thereby improving the reliability of the results.

### 2.2. Participants

A total of 474 respondents participated in the study. However, not all participants provided associations for all 30 cues. To ensure sufficient data for analysis, it was decided to exclude participants who provided fewer than 10 associations from the analyzed data. As a result, the data of 55 participants were discarded. Additionally, some questionnaires were excluded because the participants were under 16 years old<sup>4</sup>. Ultimately, the responses of 389 participants were included in the analysis. Participants' demographic distribution is detailed below.

<sup>2</sup> The categories for place of residence by population size were established based on the Law on Administrative Units and Their Boundaries of the Republic of Lithuania. Available online: <https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.5911/asr?fbclid=IwAR3MhDzZkyilAgXx2jMiSdNaDAP-w-6CuJPJVenC7LF2vR6oKnGTjwuUkE>

<sup>3</sup> Online access: <https://clarin.vdu.lt/xmlui/handle/20.500.11821/41>

<sup>4</sup> According to the General Data Protection Regulation (GDPR), Article 8, which addresses the conditions for consent of children in relation to information society services.

The participants were distributed by gender as follows: 279 women, 108 men and 2 individuals who selected the option “Other”. The responses of the latter group are included in the study; however, due to the small sample size, this group is not analyzed in terms of gender.

Participants of various ages agreed to take part in the study. As shown in Table 1, the majority of participants are aged between 21 and 30 years. The youngest participant is 16 years old and the oldest is 73 years old. The average age of the participants is 36.24 years, and the median is 31 years. It is important to note that participants were grouped by age only for data summarization purposes, these groups were not used in statistical analyses.

**Table 1.** Distribution of participants by age

| Age group   | Number of participants |
|-------------|------------------------|
| 16-20       | 28                     |
| 21-30       | 161                    |
| 31-40       | 58                     |
| 41-50       | 61                     |
| 51-60       | 61                     |
| 61-70       | 14                     |
| 71-80       | 5                      |
| No response | 1                      |

Since the study aimed to determine whether the participants’ place of residence affects word associations, they were asked to specify the size of their settlement and its region. Most participants (183) lived in large cities (Vilnius, Kaunas, Klaipėda), while 117 were from cities with populations over 3,000. There were 55 participants from villages (under 500 residents), and 34 from towns (populations between 500 and 3,000). By region, the distribution of participants was as follows:

- Dzūkija: 171 participants
- Aukštaitija: 110 participants
- Suvalkija: 66 participants
- Žemaitija: 31 participants
- Unspecified region: 11 participants

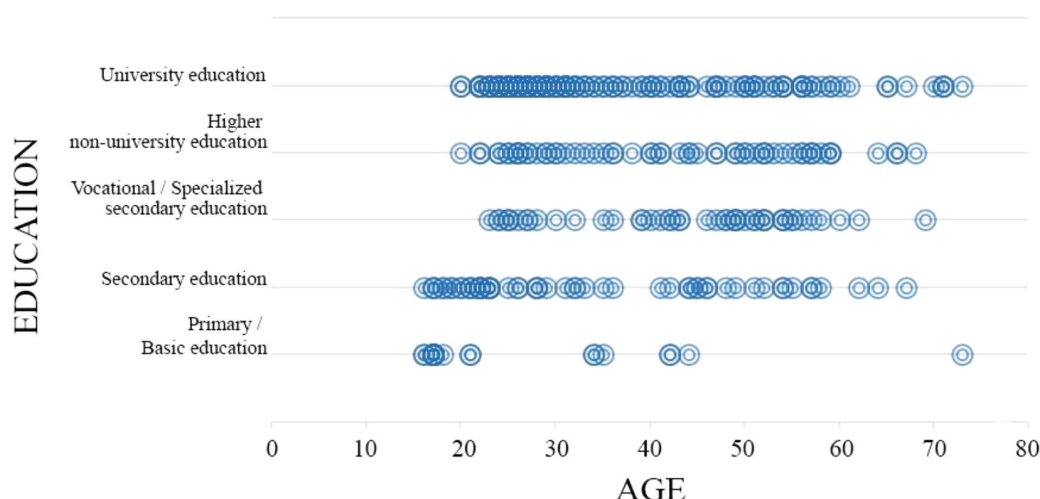
Participants were also asked about their educational background. The distribution of respondents in this regard is shown in Table 2. Most participants had higher education, while the least had primary/basic education.

**Table 2.** Distribution of participants by education

| Educational attainment | Primary / Basic education | Secondary education | Vocational / Specialized secondary education | Higher non-university education | University education | Total |
|------------------------|---------------------------|---------------------|----------------------------------------------|---------------------------------|----------------------|-------|
| Number of participants | 24                        | 58                  | 47                                           | 61                              | 199                  | 389   |

As previously mentioned, many word association studies focus on relatively homogeneous groups in terms of education. Figure 1 illustrates the distribution of age and education among participants in this study, highlighting a well-balanced representation across both variables.





**Figure 1.** Distribution of participants by age and education

It is worth noting that the data were collected online, making the sample convenient, and the groups of participants are not of equal size. However, the aim in the study was for each group to have at least 30 participants to apply the regression analysis method, which was achieved except for the group with primary/basic education. This group of participants is not analyzed in detail concerning education.

### 2.3. Data processing of the study

It was decided to follow the methodology for word association research described by Fitzpatrick et al. (2015); therefore, the data processing stage adhered to their procedure.

First, spelling errors in responses were corrected only when the respondent's intention was clear. For example, the response *bejegiskumas* to the cue *BAIMĖ* was corrected to *bejėgiškumas*, and *dkausmas* was corrected to *skausmas*. However, if a spelling error was possible but the response was an existing word in Lithuanian, the potential mistake was not corrected. For example, the response *lyga* to the cue *BAIMĖ* is a real word in Lithuanian, therefore was left uncorrected, even though many associations for this cue were *liga*, suggesting that the participant may have intended this word.

Although most responses were one-word (91% of the collected data), participants sometimes provided associations consisting of two or more words. When a response included multiple words separated by commas, only the first word was retained as the association. For example, the response *Agresija, panieka, įžeidimas* to the cue *PYKTIS* was shortened to *agresija*.

Lemmatization was considered but ultimately not applied in this study. Researchers take different approaches to lemmatization: some lemmatize responses to their strongest association (Gollan et al. 2006), others standardize only specific forms, such as plurals (Hirsh and Tree 2001). However, most studies do not provide detailed information on their lemmatization process. In this study, lemmatization was not performed due to the complexity of Lithuanian morphology. Several cases illustrate the challenges:

- The cue *MEILĖ* triggered associations *besąlyginė* and *besąlygiška*. Although these words have different suffixes, their meanings are identical.
- The cue *DRAUGYSTĖ* triggered *apkabinimas* and *apsikabinimas*. The second word is a reflexive noun, yet the meaning difference is minimal, if any.
- The cue *KALTĖ* triggered *bloga* (adj.) and *blogai* (adv.), which can sometimes be used interchangeably in Lithuanian.

These examples show that lemmatization decisions would have been highly subjective and could lead to inconsistencies, making it impractical for this study.

## 2.4. Evaluation of data according to stereotypy and uniqueness

Stereotypy in word associations refers to how frequently the same association is given to a particular cue word. Researchers assess word association stereotypy using different methods, which, according to Fitzpatrick et al. (2015: 33), generally fall into two categories: weighted and non-weighted. In this study, a weighted approach was applied, meaning that each association received as many points as the number of participants who provided the same response.

To gain additional insights, some studies also examine the uniqueness of associations (Hirsh and Tree 2001), as these two measures can complement each other. Therefore, this study also assessed uniqueness by assigning one point to associations that were given by only a single participant. Both measures were used to determine whether certain participant groups were more inclined to provide stereotypical or unique associations.

To perform the analysis, the following steps were taken:

1. All associations provided for a cue were counted and summarized into a norm list, which included each association along with its frequency. For example, the associations elicited by the cue *ATEITIS* (as shown in Figure 2) were alphabetically arranged, and their frequency was summarized as follows: *abejonės* (1), *abstraktas* (1), *anūkai* (1), *ateis* (1), *ateiviai* (5), *atsparumas* (1) and so on.

| Ateitis    |
|------------|
| Abejonės   |
| Abstraktas |
| Anūkai     |
| Ateis      |
| Ateiviai   |
| Ateiviai   |
| ateiviai   |
| Ateiviai   |
| Ateiviai   |
| Atsparumas |

**Figure 2.** Some of the word associations elicited by the cue *ATEITIS*

2. Stereotypy scores were calculated by assigning each association as many points as the number of participants who provided the same response. This means that a response received a score of 1 if only one participant provided it, including the participant themselves. Thus, the stereotypy points assigned to the associations shown in Figure 2 were as follows: *abejonės* (1) one point, *abstraktas* (1) one point, *anūkai* (1) one point, *ateis* (1) one point, *ateiviai* (5) five points, *atsparumas* (1) one point. To ensure comparability, each participant's total stereotypy points were divided by the number of associations they provided. This step was necessary because participants contributed varying numbers of responses. Without this adjustment, those who provided fewer associations (e.g., 10) would receive disproportionately lower stereotypy scores than those who provided more (e.g., 30). The result was stereotypy score for each participant.



3. Uniqueness of word associations was evaluated as follows: one point was awarded for each unique association that no one else provided for that cue. Therefore, the associations *abejonės* (1), *abstraktas* (1), *anūkai* (1), *ateis* (1) and *atsparumas* (1) indicated in Figure 2 were each awarded one uniqueness point, while the association *ateiviai* (5) was awarded zero uniqueness points. Just as with stereotypy, each participant's uniqueness points were summed and divided by the number of word associations provided by the same participant, yielding a uniqueness score for that participant's associations.

After completing these steps, each participant received stereotypy and uniqueness scores. These scores can be interpreted directly: the higher the participant's stereotypy score, the greater their tendency to provide stereotypical associations, while the higher the uniqueness score, the more unique word associations they gave. These scores were used in regression analysis, conducted in R Studio (version 4.0.2) software.

### 3. Results

Two multiple regression analyses were conducted: one with stereotypy scores and the other with uniqueness scores.

After calculating each participant's overall stereotypy score using the procedure described in Section 2.4, a regression analysis was performed, where these scores served as the dependent variable. In the initial model, there were five independent variables: gender, age, settlement size, region and education. Statistically insignificant independent variables were eliminated one by one during the analysis<sup>5</sup>. In the final model for stereotypy, settlement size, region and education were found to be insignificant. Table 3 presents the final model, which shows that gender and age significantly influenced the stereotypy of word associations.

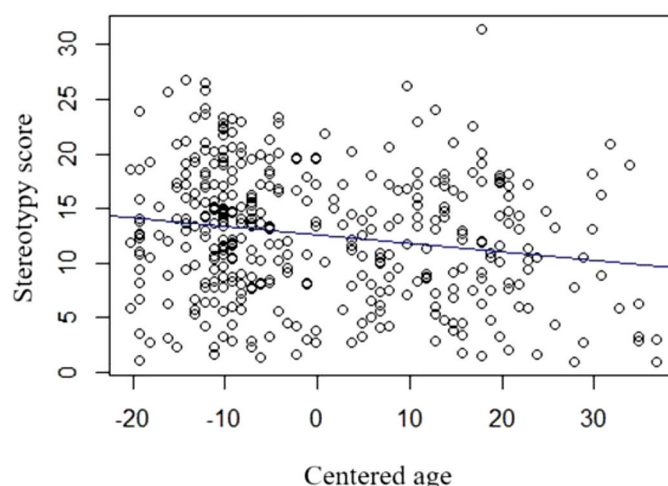
**Table 3.** Stereotypy model

| Stereotypy of word associations                       |             |                |         |         |
|-------------------------------------------------------|-------------|----------------|---------|---------|
|                                                       | Coefficient | Standard Error | t-value | p-value |
| Intercept                                             | 13,236      | 0,347          | 38,145  | < 0,001 |
| Gender (male)                                         | -2,256      | 0,675          | -3,341  | < 0,001 |
| Centered age <sup>6</sup>                             | -0,099      | 0,0218         | -4,547  | < 0,001 |
| F(3, 384) = 9,356; $p < 0,001$ R <sup>2</sup> = 0,068 |             |                |         |         |

The influence of age on the stereotypy of word associations is significant; however, the relationship between these variables is weak (the age coefficient in the model is -0.099). This can also be seen in Figure 3, where the line summarizing the relationship between the specified variables has only a slight downward slope. The relationship between the variables is negative: the older the participants, the lower their stereotypy scores for associations. This means that older participants provided fewer stereotypical associations.

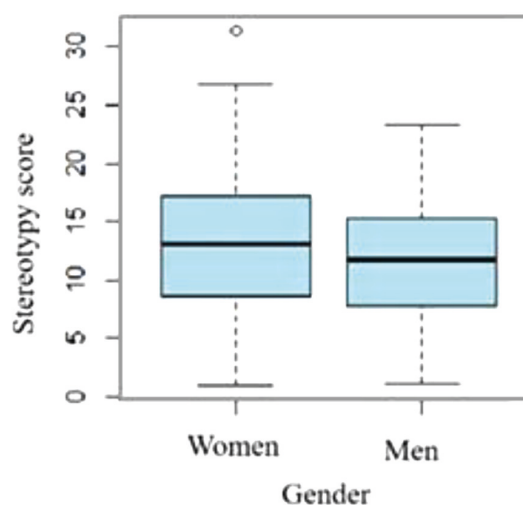
<sup>5</sup> This is a standard practice. As Čekanavičius and Murauskas (2014: 34) explain, insignificant variables are typically excluded from the final model unless special circumstances warrant their inclusion. Since no such circumstances applied here, backward stepwise selection was used (Levshina, 2015: 152).

<sup>6</sup> Data centering is a statistical data transformation procedure in which the mean of all values of a variable is subtracted from those values (Field 2009: 786). If the age variable is not centered, the intercept would indicate the predicted stereotypy score for a participant who is 0 years old.



**Figure 3.** The relationship between stereotype scores and centered participant age

The analysis revealed a stronger relationship between the stereotype score and gender: in the stereotype model, the coefficient for males is -2.256. This indicates that men provided stereotypical word associations less frequently than women. The relationship between the stereotype score and gender is illustrated in Figure 4.



**Figure 4.** The relationship between stereotype scores and participants' gender

The figure shows that the stereotype of word associations differed systematically between men and women. The highest stereotype scores for women exceeded those for men, with the first quartile, third quartile and median values all positioned higher for women. This suggests that women were more likely to provide stereotypical word associations than men.

While these findings highlight important influences on word associations, the regression model yielded a low  $R^2$  value (0.068), indicating that most variance in stereotype remains unexplained. However, there are several reasons why this model should not be dismissed entirely. First, language – including word associations – is inherently variable and influenced by numerous factors, making it challenging

to fully capture with a single model. Second, the primary aim of this study was to identify whether and which demographic factors might play a role in shaping word associations and to integrate them into a single, comprehensive model rather than analyzing them in isolation. Despite its limitations, the model successfully reveals systematic relationships between demographic factors and word association patterns, supporting findings from previous research.

## 4. Discussion

This study demonstrated that demographic factors do influence the stereotypy of word associations. The regression analysis showed that younger participants produced more stereotypical associations than older participants, highlighting age as an important factor. These results align with those found by Tresselt and Mayzner (1964) and Dubossarsky et al. (2017), which indicated that older participants' associations were less stereotypical. Dubossarsky et al. (2017) further demonstrated that associative networks evolve over the lifespan in a U-shaped pattern: during early life, associations become increasingly structured and conventionalized, while in later life, this pattern reverses, leading to greater variability in word associations. This suggests that the observed decrease in stereotypy with age may be part of a broader process in which older individuals develop a more diverse and less rigidly structured mental lexicon. While the present study does not track changes across the entire lifespan and cannot directly observe the U-shaped trajectory identified by Dubossarsky et al. (2017), its findings align with the trend of decreasing stereotypy in adulthood. This suggests that even without capturing the full developmental curve, demographic factors such as age systematically shape word association patterns.

However, the results of this study differ from those of Hirsh and Tree (2001), who found that younger participants provided a wider variety of associations than older participants. One possible reason for this discrepancy is that in the Hirsh and Tree (2001) study, most participants in the younger group had higher education, while only one participant in the older group had such an education. In contrast, the present study included a more balanced distribution of education across age groups, with older participants also having higher education (as noted in 2.2.). This suggests that the relationship between age and stereotypy in word associations may be more robust when education is more evenly accounted for. Additionally, while Hirsh and Tree (2001) examined two distinct age groups, this study treated age as a continuous variable, allowing for a more detailed perspective on how stereotypy changes gradually over the lifespan, rather than being limited to comparisons between predefined groups. Future research could further explore the interplay between age and education using models that assess their combined influence rather than examining them in isolation.

This study also showed that gender influences the stereotypy of word associations. The associations provided by men were less stereotypical than those provided by women. A similar trend was observed in the study by Jenkins and Palermo (1965). While De Deyne et al. (2023) did not examine stereotypy, their findings suggest that men and women differ in how they conceptualize word meanings, which may, in turn, contribute to gender-related differences in word associations. Since there are currently few studies examining the influence of gender on word associations, making broader generalizations is difficult; however, these findings may encourage further exploration of this demographic factor in word association research.

While a suitable model for stereotypy scores was successfully developed, a similar model for uniqueness scores could not be established. A regression analysis was performed with uniqueness scores as the dependent variable and gender, age, size of residence, region and education as independent variables. As the final model did not meet key statistical assumptions, it is not presented in further detail.

The results of this study hold value not only for researchers investigating the factors that influence word associations but also for those exploring word associations on a broader scale. While the findings do not deny that many associations recur across speakers, they indicate that such regularities should not be regarded as universal norms. Instead, the results suggest that word associations are influenced by demographic factors such as age and gender, making them more individual than previously assumed. Consequently, norm lists may be informative but can also oversimplify the complexity of word associations. At the same time, they remain valuable for studies on word meaning, provided their modeling is refined. Fitzpatrick et al. (2015: 33) suggest that study-specific norm lists, tailored to the unique characteristics of the participants in each research context, may often be more appropriate. Tailoring these lists to the specific sample under study could yield more accurate and relevant conclusions. Additionally, norm lists could be enriched by incorporating more detailed information beyond cue-association pairs. For instance, Warriner et al. (2013) compiled a norm list of affective word meanings, incorporating variations by gender, age and education level. Such an approach could offer deeper insights into the factors influencing word associations. The present study suggests that this approach could serve as a valuable direction for future word association research.

It is important to acknowledge that the conclusions of this study are not definitive. The list of cues could be expanded to include words from other parts of speech, which would provide a better understanding of whether participants' associations differ in response to cues with different linguistic characteristics. Another limitation is that this study analyzed the influence of demographic factors on word associations without examining the influence of the cues. It would be appropriate to apply a method that measures both participant and cue effects on associations within a single model, which is likely to enhance our understanding of how these factors interact.

## References

- Aitchison, J. 2003. *Words in the Mind* (3rd ed.). Oxford: Blackwell.
- Akelaitienė, G. (2007). Gyvūnų pavadinimai ir kalbinės asociacijos. *Žmogus ir žodis*, 9(1), 65-69.
- Bandera, L., Della Sala, S., Laiacina, M., Luzzatti, C., Spinnler, H. 1991. Generative associative naming in dementia of Alzheimer's type. *Neuropsychologia*, 29(4), 291-304.
- Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., Tummolini, L. 2017. The challenge of abstract concepts. *Psychological Bulletin*, 143(3), 263.
- Cremer, M., Dingshoff, D., de Beer, M., Schoonen, R. 2011. Do word associations assess word knowledge? A comparison of L1 and L2, child and adult word associations. *International Journal of Bilingualism*, 15(2), 187-204.
- Čekanavičius, V., Murauskas, G. 2014. *Taikomoji regresinė analizė socialiniuose tyrimuose*. Vilnius: Vilniaus universiteto leidykla.
- De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., Storms, G. 2019. The "Small World of Words". English word association norms for over 12,000 cue words. *Behavior Research Methods*, 51(3), 987-1006.
- De Deyne, S., Storms, G. 2008. Word associations: Network and semantic properties. *Behavior Research Methods*, 40(1), 213-231.
- De Deyne, S., Warner, S., Perfors, A. 2023. Common words, uncommon meanings: Evidence for widespread gender differences in word meaning. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45, No. 45).
- de Groot, A. M. 1989. Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 824.

- Dubossarsky, H., De Deyne, S., Hills, T. T. 2017. Quantifying the structure of free association networks across the life span. *Developmental Psychology*, 53(8), 1560.
- Field, A. 2009. *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications Ltd.
- Fitzpatrick, T. 2007. Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics*, 17(3), 319-331.
- Fitzpatrick, T. 2009. Word association profiles in a first and second language: Puzzles and problems. In T. Fitzpatrick and A. Barfield (eds.) *Lexical processing in second language learners* (pp. 38-52). Bristol: Multilingual Matters.
- Fitzpatrick, T., Playfoot, D., Wray, A., Wright, M. J. 2015. Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics*, 36(1), 23-50.
- Fitzpatrick, T., Thwaites, P. 2020. Word association research and the L2 lexicon. *Language Teaching*, 53(3), 237-274.
- Gollan, T. H., Salmon, D. P., Paxton, J. L. 2006. Word association in early Alzheimer's disease. *Brain and Language*, 99(3), 289-303.
- Higginbotham, G. 2010. Individual learner profiles from word association tests: The effect of word frequency. *System*, 38(3), 379-390.
- Hirsh, K. W., Tree, J. J. 2001. Word association norms for two cohorts of British adults. *Journal of Neurolinguistics*, 14(1), 1-44.
- Jenkins, J. J., Palermo, D. S. 1965. Further data on changes in word-association norms. *Journal of Personality and Social Psychology*, 1(4), 303.
- Jung, C. G. 1910. The association method. *The American Journal of Psychology*, 21(2), 219-269.
- Karimkhani, F., Rahmani, H., Zare, A., Saheb Nassagh, R., Aghakasiri, K. 2021. Tarvaje: Word association norms for Persian words. *Journal of Psycholinguistic Research*, 50(4), 863-882.
- Levshina, N. 2015. *How to do linguistics with R*. Netherlands: John Benjamins Publishing Company.
- Lin, S. Y., Chen, H. C., Chang, T. H., Lee, W. E., Sung, Y. T. 2019. CLAD: A corpus-derived Chinese lexical association database. *Behavior Research Methods*, 51(5), 2310-2336.
- Mattheoudakis, M. 2011. A word-association study in Greek and the concept of the syntagmatic-paradigmatic shift. *Journal of Greek Linguistics*, 11(2), 167-197.
- Meara, P. 1983. Word associations in a foreign language. *Nottingham Linguistics Circular*, 11(2), 29-38.
- Nelson, D. L., McEvoy, C. L., Schreiber, T. A. 2004. The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3), 402-407.
- Nissen, H. B., Henriksen, B. 2006. Word class influence on word association test results. *International Journal of Applied Linguistics*, 16(3), 389-408.
- Palermo, D. S., Jenkins, J. J. 1965. Sex differences in word associations. *Journal of General Psychology*, 72, 77.
- Papaurėlytė, S. (2011). Žodžio medis asociacijų laukas lietuvių kalbos pasaulėvaizdyje. *Acta humanitarica universitatis Saulensis*, 13, 405-416.
- Papaurėlytė, S. (2012). Žodžio žodis asociacijų laukas lietuvių kalbos pasaulėvaizdyje. *Žmogus ir žodis*, 14(1), 103-109.
- Papaurėlytė, S. (2014). Žodžio kaimas asociacijų laukas lietuvių kalbos pasaulėvaizdyje. *Acta humanitarica universitatis Saulensis*, 19, 351-362.
- Petchkovsky, L., Petchkovsky, M., Morris, P., Dickson, P., Montgomery, D., Dwyer, J., Burnett, P. 2013. fMRI responses to Jung's Word Association Test: implications for theory, treatment and research. *Journal of Analytical Psychology*, 58(3), 409-431.
- Playfoot, D., Balint, T., Pandya, V., Parkes, A., Peters, M., Richards, S. 2018. Are word association responses really the first words that come to mind? *Applied Linguistics*, 39(5), 607-624.
- Schmitt, N. 2010. *Researching vocabulary: a vocabulary research manual*. New York, NY: Palgrave Macmillan.
- Steponavičienė, S. 1986. *Lietuvių kalbos žodinių asociacijų žodynas*. Vilnius: Mokslo.

- Stoet, G. 2010. PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096-1104.
- Stoet, G. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31.
- Tresselt, M. E., Leeds, D. S., Mayzner Jr, M. S. 1955. The Kent-Rosanoff word association: A comparison of sex differences in response frequencies. *The Journal of Genetic Psychology*, 87(1), 149-153.
- Tresselt, M. E., Mayzner, M. S. 1964. The Kent-Rosanoff word association: Word association norms as a function of age. *Psychonomic Science*, 1(1), 65-66.
- Vilkaitė-Lozdienė, L. 2019a. First steps towards the Lithuanian word association database. *Taikomoji kalbotyra*, 12, 226-258. doi: 10.15388/TK.2019.17238.
- Vilkaitė-Lozdienė, L. 2019b. Lexical associations Does the response depend on the cue's morphology? Presented at the *Kalbos ir žmonės: komunikacija daugiakalbiame pasaulyje*, Vilnius.
- Vitrano, D., Altarriba, J., Leblebici-Basar, D. 2021. Revisiting Mednick's (1962) theory of creativity with a composite measure of creativity: The effect of stimulus type on word association production. *The Journal of Creative Behavior*, 55(4), 925-936.
- Warriner, A. B., Kuperman, V., Brysbaert, M. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191-1207.
- Wettler, M., Rapp, R., Sedlmeier, P. 2005. Free word associations correspond to contiguities between words in texts. *Journal of Quantitative Linguistics*, 12(2-3), 111-122.
- Wolter, B. 2001. Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, 23(1), 41-69.
- Zeelenberg, R., Pecher, D., Shiffrin, R. M., Raaijmakers, J. G. 2003. Semantic context effects and priming in word association. *Psychonomic Bulletin & Review*, 10(3), 653-660.

## Data sources

- Dadurkevičius, V. 2020. Wordlist of Lemmas from the Joint Corpus of Lithuanian, CLARIN-LT digital library in the Republic of Lithuania, <https://clarin.vdu.lt/xmlui/handle/20.500.11821/41>.

Submitted March 2025

Accepted June 2025



## Appendix. Cue words

|             |              |            |          |
|-------------|--------------|------------|----------|
| Laikas      | Meilė        | Džiaugsmas | Vienybė  |
| Šeima       | Laimė        | Vaikystė   | Kančia   |
| Pabaiga     | Garbė        | Nerimas    | Santarvė |
| Laisvė      | Taika        | Gėda       | Lygybė   |
| Tiesa       | Laisvalaikis | Draugystė  | Pavydas  |
| Amžius      | Kaltė        | Pyktis     | Dabartis |
| Ateitis     | Akimirka     | Senovė     |          |
| Teisingumas | Baimė        | Kasdienybė |          |