DOI: https://doi.org/10.18454/RULB.2021.25.1.30

НОРМАТИВЫ СЕМАНТИЧЕСКИХ КАТЕГОРИЙ НА АНГЛИЙСКОМ ЯЗЫКЕ ДЛЯ НОСИТЕЛЕЙ РУССКОГО ЯЗЫКА

Научная статья

Марченко О.П. *
ORCID: 0000-0002-5716-6744,
Московский государственный психолого-педагогический университет, Москва, Россия

* Корреспондирующий автор (olga.p.marchenko[at]gmail.com)

Аннотация

Цель исследования - сравнить сходство частоты называния слов в семантических категориях на родном (L1) и иностранном (L2) языках при разных уровнях языковой компетенции, и разработать категориальные нормативы для английского как иностранного. В выборку вошли 120 русскоговорящих студентов с высоким, средним и низким уровнем владения L2. Участников попросили перечислить за 30 секунд как можно больше слов, относящихся к категории. Рассчитывалась частота каждого слова в категории для русского и английского языков для всей выборки и отдельно для группы с высоким, средним и низким уровнем владения языком. Для измерения схожести частотных распределений слов для каждой категории между русским и английским языками использовались коэффициенты близости Хеллингера (HA). Уровень сходства между частотными распределениями на родном и иностранном языках был сильнее в группе с высоким уровнем владения вторым языком, по сравнению с группой с низким уровнем владения. Это показывает, что уровень владения языком важен для того, чтобы дать человеку возможность выразить себя. Стабильные различия между английским языком как L2 по сравнению с американскими нормами подтверждают необходимость использования для тех индивидов, для которых английский язык является иностранным, нормативов семантических категорий, разработанных специально для них. Высокий уровень надежности позволяет использовать эти частотные показатели 45 семантических категорий как нормативные для носителей русского языка, изучающих английский язык. Полученные категориальные нормативы для английского языка как иностранного можно использовать для диагностики когнитивных нарушений и исследований категоризации в L2 на русскоговорящей выборке, поскольку показатели частоты называния слов могут повлиять на результаты.

Ключевые слова: уровень владения языком, родной язык, иностранный язык, категориальные нормативы, частота называния.

SEMANTIC CATEGORY NORMS OF ENGLISH LANGUAGE FOR NATIVE RUSSIAN SPEAKERS

Research article

Marchenko O.P. *
ORCID: 0000-0002-5716-6744,
Moscow State University of Psychology and Education, Moscow, Russia

* Corresponding author (olga.p.marchenko[at]gmail.com)

Abstract

This study aimed to compare the similarity (proximity) of category norms in native (L1) and foreign languages (L2) in different levels of language proficiency and to develop category norms of the English language as L2 for Russian native speakers. The sample included 120 students with high, average, and low levels of L2 proficiency. Participants were instructed to write as many words which come to mind from a category as they could in 30 seconds. The total frequency of each word in a category for Russian and English languages was determined for the whole sample and separately for the group with high, average, and low levels of language proficiency. Aiming to determine the overlap of the exemplar frequencies for each category between Russian and English in three different levels of language proficiency, Hellinger Affinity (HA) indexes were computed. The level of affinity between frequency distributions in native and foreign languages was stronger in the group with a high level of L2 proficiency compared to the group with a low level of proficiency. It shows that the level of language acquisition is important to give a person the possibility to express him/herself. Stable differences between the English language as L2 compared to North American norms support the necessity of using special category norms elaborated for L2 in studies of non-native English speakers. Strong reliability indexes for categories suggest that these frequency scores for 45 semantic categories can be accepted as normative for native Russian speakers learning English as L2. Category norms for English as a second language can be applied for diagnosing cognitive impairments and for studies of categorization using L2 in a Russian sample as a generation frequency variable can affect the results.

Keywords: language proficiency, L1, L2, category norms, exemplar generation frequency.

Introduction

Considering the fact that language and culture are interdependent, the problem of cross-cultural differences in categorization raises the question of how language acquisition influences the category structure. It has been found previously that there is a great impact of the native language's conceptual organization on category generation in the foreign language for culture-specific categories (e.g., *Kind of Food, Vehicles*) [2]. One can suggest that the level of language competence along with cultural background can affect the category structure. Thus, the present study was aimed to compare the similarity of the structure of a category in the first and second languages at different levels of foreign language proficiency. The method of naming items from a particular semantic category enables to determine the category content and the level of dominance of items in a category (category frequency of words). The divergence in the organization of a category in the first (L1) and second

language (L2) can be measured through the overlap between the generation frequencies of words named in this category in the first and second languages. We hypothesized that the level of affinity between frequency distributions in native and foreign languages will be stronger in the group with a high level of L2 proficiency compared to the group with an average and low level of L2 proficiency. Another goal of the study was to test whether the affinity between L2 norms and North American norms is getting stronger if the level of L2 proficiency grows. If the affinity with North American norms is consistent with the level of proficiency of L2, it will show that category norms depend mainly on the level of language proficiency and are affected by the involvement in the culture, the language represents. In case the affinity with North American norms will not be consistent with the level of L2 proficiency, one can suggest that at least in academic bilinguals, native language culture uniquely affects category norms in L2. If this suggestion meets the reality, then L2 category norms cannot be replaced by native L1 norms developed in the native speakers' population. Thus, in this case, another purpose of the research was to develop category dominance norms for English as an L2 language in Russian speakers' sample.

Methods

Participants

One hundred twenty Russian speakers 18-27 years (Mean=20(1.9), 15 men) who are English language learners were recruited in the study. The sample included undergraduate students with a level of language proficiency ranging from A2 to C2 according to Common European Framework classification, and from 2 to 5 according to the Russian university grade system. Part of the participants were undergraduate students majoring in Linguistics. Participants were divided into three groups according to university scores, the Common European Framework classification, experience of using language in everyday life, and verbal fluency. The sample included 40 students with a high level of second language proficiency (Mean=20(1.9), 38 women), 40 students with an average level of second language proficiency (Mean=20(2.05), 30 women), and 40 students with a low level of second language proficiency (Mean=20(1.2), 37 women).

Procedure

Participants were instructed to write as many words which come to mind from a category as they could in 30 seconds. The task was performed for 45 semantic categories (Vegetables, Birds, Trees, Musical Instruments, Toys, etc.).

Each student carried out the task in Russian (L1) and English language (L2). The list of categories was randomized. The total frequency of each word in a category for Russian and English language and the frequency item was named first in the list were determined for the whole sample and separately for the group with high, average, and low levels of language proficiency. Aiming to determine the overlap of the exemplar frequencies for each category between Russian and English in three different levels of language proficiency, we computed Hellinger Affinity (HA) indexes [6]. The value of HA indexes varies from 0 to 1, where 1 means the maximal match of two frequency lists. Friedman's χ^2r test, Wilcoxon T-test, and Student T-test were used to compare HA scores between L1, L2, and North American norms at different levels of proficiency. Pearson's χ^2-test with Yates' correction for continuity was applied to compare the total generation frequency of each word between the first and second language in three levels of proficiency. For comparing the ratios of exemplars, which had significant differences ($p<0.05$) between L2 and NA in different levels of language proficiency, the two-tailed Pearson χ^2-test was used. North American category dominance norms for 35 categories (NA) were taken from Van Overschelde et al. [4], [5]. HA indexes for category dominance norms between samples from different regions for the Russian language were taken from Marchenko et al. [3]. For statistical analyses, IBM SPSS 25 and Microsoft Excel were used.

Results

For words in 45 categories, the total frequency and frequency item was named first in the list were calculated in both languages. The reliability Spearman-Brown indexes for categories in L2 were strong (m=0.99, see supplementary material: Table 1) shows that these category dominance values can be accepted as normative for native Russian speakers learning English as a second language. Category dominance values for words named at least two times are provided in Supplementary materials for this paper (see Supplementary material: Table 2).

The HA was counted for frequency distributions of Russian and English languages, as well as for English and NA norms separately for samples of high, average, and low levels of language proficiency (see Table 1). HA indexes were measured between two randomly divided samples of second language data (Split-half HA) with the purpose to present zero affinity as well.

Table 1 – HA indexes in three levels of language proficiency

	Median	Mean	SD
High level of L2 proficiency -NA	.74	.74	.13
Average level of L2 proficiency -NA	.73	.74	.14
Low level of L2 proficiency -NA	.73	.72	.13
High level of L2 proficiency -L1	.84	.78	.15
Average level of L2 proficiency -L1	.83	.80	.13
Low level of L2 proficiency -L1	.79	.77	.13
Split-half HA	.92	.90	.07

HA of L2 and NA norms have no significant differences between high, average, and low level of proficiency samples (S=1,984, p=.371). Such a result shows the stability of cross-cultural differences. The affinity of category dominance distributions of L2 and L1 were significantly stronger than the affinity of L2 and North American norms at all three levels of

proficiency (Z=-2,546, p<.05 for a high level of L2 proficiency, Z=-3,252, p<.001 for an average level of L2 proficiency, Z=-3,108, p<.01 for a low level of L2 proficiency). It shows the cultural specificity of the data.

The level of affinity between the first and second language was lower than the split-half affinity for L2 (in average Z=-5.82, ps<.001) or than the intracultural level of affinity for L1 (in average Z=-5.84, ps<.001).

There were differences in the number of words named in L2 compared to L1 (t (44) =8.867, p<.001 for a high level of L2 proficiency, t (44) =11.642, p<.001 for an average level of L2 proficiency, t (44) =14.332, p<.001 for a low level of L2 proficiency). The difference in fluency between L1 and L2 was stronger in the group with a low level of language proficiency compared to an average level group (Z=-2,241, p<.05) and in the group with an average level compared to the high level group (Z=-2,542, p<.05). The difference in fluency between L2 and L1 was 54 words per category in the group with a high level of proficiency, 77 words per category in the group with an average level of proficiency, and 97 words per category in the group with a low level of proficiency. Such dissimilarity in word fluency in L1 and L2 can affect the frequency of words in these two cases.

The level of affinity between frequency distributions in L1 and L2 was stronger in the group with a high level of second language proficiency compared to the group with a low level of L2 proficiency (Z=-2,092, p<.05) and to the group with an average level of second language proficiency (Z=-3,359, p<.001). It shows that raising the level of language acquisition enables a person to express him/herself. Thus, the development of language proficiency leads to some change in category dominance scores.

When a person does not interact with the objects using a second language, he will presumably translate concepts that are familiar to him from the first language. Thus, individuals using L2 may be influenced by the conceptual knowledge of the native language. That is why the level of proximity of frequency distributions of words named in the first and second languages was stronger in the sample with a high level of second language proficiency (due to the better developed vocabulary in L2).

The level of dominance for L2 may depend on vocabulary development. L2 lexicon limitations affected the total frequency of words and the frequency pattern of the category. When a subject uses English, he/she in some cases begins to name culture-specific concepts (e.g., English names in *Girls` and Boys` First Names*). Nevertheless, the frequency of these culture-specific concepts was lower in Russian speakers sample using L2 compared to NA in all levels of language proficiency. Additional analyses of item frequencies showed that the same concepts in all three levels of L2 proficiency had a similar pattern of differences. For example, Russian native speakers named *inches* in the category *Units of distance* less frequently than North American subjects ($\chi2$=80.65, df=1, p<.001 for a high level of L2 proficiency, $\chi2$= 63.03, df=1, p<.001 for an average level of L2 proficiency, $\chi2$= 101.47, df=1, p<.001 for a low level of L2 proficiency). A similar result was shown for the word *yard* ($\chi2$=50.02, df=1, p<.001 for a high level of L2 proficiency, $\chi2$= 44.61, df=1, p<.001 for an average level of L2 proficiency, $\chi2$= 50.02, df=1, p<.001 for a low level of L2 proficiency). There were crucial differences in naming the words *mom* and *mother*. While NA used both words with similar frequency, Russian subjects named *the mother* more frequently in L2 ($\chi2$=38.48 df=1, for a high level of L2 proficiency, $\chi2$=43.587 for an average level of L2 proficiency, $\chi2$=34.922 for a low level of L2 proficiency). An analogous result was evident for *father* and *dad*. While in the NA sample these formal and informal words had similar frequency, the Russian sample significantly more frequently preferred to name *the father* in L2 ($\chi2$= 35.52, df=1, p<.001 for a high level of L2 proficiency, $\chi2$= 40.47, df=1, p<.001 for an average level of L2 proficiency, $\chi2$= 33.15, df=1, p<.001 for a low level of L2 proficiency). According to these data, one can suggest that Russian subjects use formal terms and it can be treated as a sign that they do not use L2 in everyday life. Some words were named more frequently in the Russian sample compared to NA norms. For example, the word *armchair* in category *Furniture* was used only in Russian sample ($\chi2$=77.92, df=1, p<.001 for a high level of L2 proficiency, $\chi2$= 68.50, df=1, p<.001 for an average level of L2 proficiency, $\chi2$= 43.73, df=1, p<.001 for a low level of L2 proficiency). The stable pattern of differences in samples of all three levels of language proficiency supports the reliability of culture-specific differences of category norms. Significant differences in category dominance compared to NA norms appeared for words with high frequency as well as for words with low frequency. In general, there were more words with the same level of dominance on L2 as in NA norms than words with significantly different levels (Table 2), which resulted in acceptable HA levels. If someone will use these ratings for selecting words for experimental studies, it will be possible to find stimuli that have a similar level of dominance in both cultures. Nevertheless, strong cross-cultural differences showed in at least one-third of the words (Table 2).

Table 2 – Number of words with significant and nonsignificant differences between L2 and North American norms in the three levels of second language proficiency

	significant (p<.05)	nonsignificant (p≥.05)
High level of L2 proficiency	626	1384
Average level of L2 proficiency	607	1432
Low level of L2 proficiency	610	1194

There were more words with significantly different levels of dominance from NA in the group with a low level of language proficiency than in the group with high ($\chi2$= 3.093, df=1, p<.05) and an average level of language proficiency ($\chi2$= 7.235, df=1, p<.01). The numbers of words with different frequencies were not significantly different in groups with a high and average level of language proficiency ($\chi2$= 3.093, df=1, p=.078).

When a person used the native language, he or she used informal words more frequently than they did in L2. For example, the word *papa* (dad) was used more frequently in Russian language instead of *otec* (father) and the result was reversed for L2 ($\chi2$= 27.78, df=1, p<.001 for a high level of L2 proficiency, $\chi2$= 33.76, df=1, p<.001 for an average level of L2 proficiency, $\chi2$=47.70, df=1, p<.001 for a low level of L2 proficiency). When subjects used the native language, they more frequently named *mama* (mom) instead of *mat* (mother) and when they used English, the result was the opposite ($\chi2$= 8.80, df=1, p<.01

for a high level of L2 proficiency, $\chi 2 = 35.74$, df=1, p<.001 for an average level of L2 proficiency, $\chi 2 = 45.73$, df=1, p<.001 for a low level of L2 proficiency).

Given the strong differences between Russian norms for English as a foreign language and NA norms in a category structure, it is important to use the category norms of L2 in the Russian sample for selecting words for experimental studies. As the pattern of differences was similar for all three levels of proficiency, the data can be united in the same sample.

Conclusion

Category dominance scores for words of 45 different semantic categories for English as a foreign language are presented. Strong reliability indexes imply that these frequency scores for 45 semantic categories can be accepted as normative for native Russian speakers learning English as L2. As the affinity with North American norms is not related to the level of L2 proficiency, it shows that, at least in academic bilinguals, category norms in a second language are uniquely affected by the culture of the native language and cannot be replaced by norms developed abroad. Category norms for the English language as L2 can be applied for the assessment of cognitive deficits and for studies of categorization using L2 in a Russian sample as a generation frequency variable can affect the results. Stable differences between English as L2 language and NA norms support the necessity of using local L2 category norms instead of norms developed in the country of the language. The sample of Russian speakers who use the English language in the educational environment is quite extensive; therefore, this database will be useful.

Финансирование

Эта работа была поддержана Российским фондом фундаментальных исследований, проект 20-013-00912.

Funding

This work was supported by Russian Foundation for Basic Research, grant 20-013-00912.

Конфликт интересов

Не указан.

Conflict of Interest

None declared.

Список литературы / References

1. Battig W.F. Category norms for verbal items in 56 categories: a replication and extension of the Connecticut category norms / Battig W.F., & Montague W.E. // Journal of Experimental Psychology Monograph. – 1969. – Vol. 80(3). – P. 1-45.
2. Marchenko O.P. Comparing the Generation of Words from Different Semantic Categories in Native and Foreign Languages / Marchenko O.P. // Psychology in Russia. State of Art. – 2019. – 1. – P. 163-171.
3. Marchenko O.P. Intra-cultural continuity and diversity of semantic category norm / Marchenko O.P., Bandurka T.N., & Pavlov Yu.G. // Russian Linguistic Bulletin. – 2018– 2 (14). – P. 3-7.
4. Van Overschelde J.P. Category norms: An updated and expanded version of the Battig and Montague (1969) norms / Van Overschelde J.P., Rawson K.A., & Dunlosky J. // Journal of Memory and Language. – 2004. – 50. – P. 289-335.
5. Van Overschelde, Jim & Rawson, Katherine & Dunlosky, John. Addendum to “Category norms: An updated and expanded version of the Battig and Montague (1969) norms” [Journal of Memory and Language 50 (2004) 289–335]. // Journal of Memory and Language. – 2006– 54(4). 633.
6. Yoon C. Category norms as a function of culture and age: Comparisons of item responses to 105 categories by American and Chinese adults / Yoon C., Feinberg F., Hu P. et al. // Psychology and Aging. – 2004. – Vol. 19. – P. 379-393.