

Let's keep an eye on Russian: testing sensitivity to the change in the grammatical number in somatic idioms with ruBERT

Elena Spaziani^{1,*†}, Eliana Di Palma^{1,2†}

¹Sapienza, Università di Roma, Roma, Italy

²Università degli Studi di Roma Tre, Roma, Italy

Abstract

English. In recent times, linguistic research and computational linguistics have considered the morpho-semantic properties of the grammatical number in nouns. In this context, our study aimed at using a language model to test number variation on selected Russian somatic idioms containing locally and generally marked nouns. Overall, the model's sensitivity corroborated the relevance of number in idioms, especially in instances where the semantics of number is particularly significant.

Italiano. Negli ultimi tempi la ricerca linguistica e la linguistica computazionale hanno iniziato a considerare le proprietà morfo-semantiche del numero grammaticale dei sostantivi. In tale contesto, questo studio si è proposto di utilizzare un modello linguistico per testare la variabilità del numero nei fraseologismi somatici russi con nomi dalla marcatezza locale o generale. Nel complesso, la sensibilità del modello al cambiamento di valori grammaticali ha dimostrato l'influenza del numero, soprattutto in istanze in cui questo è particolarmente significativo dal punto di vista semantico.

Keywords

grammatical number, somatic idioms, Russian BERT, computationally-aided linguistic analysis

1. Introduction

Somatic idioms are defined as phraseological units which contain lexemes referring to human or animal body parts [1, 2]. This class is found in many languages, due to the universal nature of the somatic lexicon. Somatismes are usually observed from a semantic perspective and mostly in contrastive analyses, since their meaning is regarded as a cultural-specific conceptualisation. However, to the best of our knowledge, existing inquiries, both in theoretical and computational linguistics, have not yet considered the morpho-semantic features of the grammatical number [3] as part of the overall idiomatic meaning. Furthermore, our interest in number is motivated by the fact that many somatic terms exhibit an effect of the pluralization's lexical properties. In particular, nouns denoting bipartite or complex body parts, which are naturally constituted by a plurality, show a higher frequency in the plural form in many languages. This condition is defined **local or semantic markedness** [4, 5], as the singular value becomes the locally marked member of the opposition, changing the morpho-semantic dynamics between

the two values. The goal of our work is to consider how change in the grammatical number affects the idiomatic structure of somatismes. More specifically, the questions underlying the following experiment were:

1. Does number variation have a significant impact on the idiomatic structure of Russian somatic idioms?
2. Does the type of markedness affect the probability of finding a singular or a plural value?

We hypothesized that change in number would have an effect on the overall structure of somatic idioms, especially in the cases of local markedness. In detail, we assumed that if we switched the values of a locally marked somatic term a greater impact on the idioms would be observed, compared to generally marked somatic terms.

The following sections describe how we addressed these questions, starting from an examination of related works. The new dataset of Russian somatic idioms that we constructed to test our hypotheses and its creation method are thereafter illustrated¹. Subsequently, our experiment is presented, which was conducted with ruBERT, a Russian-trained BERT, following a methodology similar to Salazar et al. [6] for acceptability, and Pedinotti et al. [7] for semantic plausibility. The results of the experiment are offered and discussed in the final sections. Ultimately, a conclusive section provides an answer to our questions, highlighting the relevance of our work and possible future expansion.

¹The dataset is available on request.

CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 – Dec 02, 2023, Venice, Italy

*Corresponding author.

†These authors contributed equally.

✉ elena.spaziani@uniroma1.it (E. Spaziani);

eliana.dipalma@uniroma1.it (E. D. Palma)

🆔 0000-0002-7140-9773 (E. Spaziani); 0000-0003-2154-2696

(E. D. Palma)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

2. Related works

Studies on idiom semantics adopting a compositional approach have claimed that the semantic analyzability or decomposability of idioms (the meaning of the constituents contributes to the overall figurative meaning) is related to their flexibility, that is their lexical or syntactical variation (see Cserép, Dobrovolskij for an overview).

In idiom variation research, number has been investigated as a flexibility dimension by Langlotz [10] and Cserép [11]. Specifically, Langlotz [10] asserts that number variation is systematical in VP-idioms or PP-idioms that feature an isomorphic, that is analyzable, semantic structure. However, this variation is believed to be prevented if the idiom contains an invariable noun or a noun which is incompatible with the global idiomatic meaning. The second study [11] adopts a corpus-driven approach and it retrieves the morpho-syntactic alternants in the noun phrase of V NP-idioms. As a result, a higher frequency in the singular value is assessed and a similar correlation between number variation and decomposability is found. Both works acknowledge the role of number in the idiomatic structure, however they are limited to the English language and no distinction based on the lexical meaning or type of markedness of the constituents has been drawn.

In concordance with the aforementioned theories on the category of number, studies on the Russian language assert that «the grammatical values of number are closely related and actively interact with lexical meaning»² [12, 24] (see also Vinogradov [13]). As observed by Ljaevskaja [14], this interaction is particularly evident in nouns that denote complex items, including complex and paired body parts. The lexical meaning of such nouns is intertwined with their numerical information as they denote a complex structure, constituted by plural members functioning or perceived as a whole. As a result, in these cases the grammemes of number convey more than just the numerical information. To assess this extra-grammatical meaning, the two values of some somatic terms have been observed, based on their use in context [14, 71-86]. The result of this detailed qualitative investigation is the distinction of two contexts: regular distributive contexts (i.e. the standard grammatical opposition: one - many); non-distributive contexts, which are mostly noun-specific, value-specific, and language-specific (As Ljaševskaja illustrates [14, 72], он шел **по колену** [sg] в грязи - он šel po kolenu v grjazi cannot be replaced with the plural он шел ***по коленам** [pl] в грязи - он šel po kolenam v grjazi, whereas in English a plural form is required: "He walked in mud up to his **knees**"). Furthermore, similar contexts lead to a classification of similar somatic terms. Apparently, глаз - glaz "eye" and ухо - ucho "ear" display

²In the original: «грамматические значения числа тесно связаны и активно взаимодействуют с лексическими значениями».

corresponding contexts, showing how different meanings associate with the numerical values. On the contrary, мозг - mozg "brain" exhibits a quite unique condition: although it denotes a single item, its pluralization features an additional meaning, absent in the singular, acting as a separate lexeme (as does the English equivalent 'brains'). Regarding the markedness, the Academic Russian Grammar [15] and Ušakov's Dictionary [16] indicate that some nouns are mostly used in the plural, including somatic terms. However, a full account for the whole somatic lexical domain, in terms of frequency and markedness, is not available as it is for other languages.

Only in recent times have the lexical properties of number been addressed in the computational field. Gromann and Declerck [17] investigate the semantic shifts created by the regular inflectional morphemes of number. The study shows the impact of these inflectional variants on morphological embeddings and their variability, on a par with derivational morphemes. Even more recently, Janzso [18] focuses on the ambiguity originating from these inflectional variants, as the plural form of some nouns carries a secondary non-grammatical meaning, absent in the singular value. Specifically, the work inquires about how ambiguity in number and gender is treated by contextual meaning representations, using four pre-trained BERT models on the disambiguation task. In the literature on idioms, computational studies focused on their identification or comprehension [19, 20, 21, 22], but to the best of our knowledge there have been no studies investigating the role of number in the construction of idioms and its interpretation.

3. Dataset

The dataset was constructed by selecting somatic idioms from two Russian phraseological dictionaries [23, 24]. Nine somatic components were considered in the selection: six terms referring to bipartite or complex body parts (**глаз** - glaz "eye", **рука** - ruka "hand/arm", **нога** - noga "foot/leg", **руба** - guba "lip", **ухо** - ucho "ear", **зуб** - zub "tooth"); three terms referring to a single body part or organ (**голова** - golova "head", **мозг** - mozg "brain", **язык** - jazyk "tongue").

In order to have a broad idea of the dominance condition [5], and therefore of the type of markedness, the relative frequency for each term was obtained from the Russian National Corpus³ (see Table 1).

These components were considered both in the **singular** and in the **plural** form. Altogether 73 somatic idioms in the **singular** and 73 somatic idioms in the **plural** were retrieved⁴. Idioms registered in both forms

³<https://ruscorpora.ru>

⁴Ideally, ten idioms were to be selected for each term. However, the actual number of idioms is due to their availability in the dic-

Table 1

Relative frequency, ipm (instances per million), drawn from the Russian National Corpus.

Noun	Freq. [sg]	Freq. [pl]
глаз - glaz "eye"	98, 31	860, 15
рука - ruka "hand/arm"	612, 6	700, 2
нога - noga "foot/leg"	136, 46	343, 33
губа - guba "lip"	20, 6	126, 31
ухо - ucho "ear"	62, 74	76, 72
зуб - zub "tooth"	16, 49	98, 81
мозг - mozg "brain"	59, 54	14, 77
голова - golova "head"	723, 34	72, 97
язык - jazyk "tongue"	303, 5	59, 15

were avoided whenever possible (по губе/по губам - po gube/po gubam "to one's taste", lit. "to one's lip/lips"). Formally, the selected instances range from a higher degree of idiomaticity (fixed idiomatic expressions as глаза на лоб полезли - glaza na lob polezli "eyes nearly popped out", lit. "eyes climbed on the forehead") to a more flexible condition (idiomatic prepositional phrases as в головах - v golovach "at the head of", lit. "in the heads").

The Russian National Corpus was queried and the contexts were derived and trimmed to simple sentences. To evaluate the impact of change in number, the original sentences were manipulated in number (**manipulated condition**). As a result, the somatic component in the original condition was inflected to the alternative numerical value in the grammatical opposition (i.e. the component in the original singular-only form was inflected in the plural form and vice versa). Subsequently, a **control condition** was created by substituting the somatic component with an alternative word. However, it should be noted that a random substitution was excluded, considering the aforementioned gradient of idiomaticity, which relates to an equivalent gradient of substitutability of the idiom's component parts [25]. Therefore, the control condition was created by considering alternative candidates for a given construction offered by the corpus⁵. From a grammatical perspective, the original numerical value was always maintained in the alternative words. On the other hand, the grammatical categories of gender and animacy were maintained as far as possible, considering their subjection to the choice of a valid semantic alternative in terms of context (e.g. the original в глазах - v glazach [m] "in the eyes" was substituted with в комнатах - v komnatach [f] "in the rooms"). Finally,

tionaries or in the Russian National Corpus. Specifically, in the singular section, ухо - ucho "ear" and зуб - zub "tooth" (9 idioms), губа - guba "lip" (2 idioms) and мозг - mozg "brain" (3 idioms) did not reach the set number. The same applies for губы - guby "lips" (6 idioms), головы - golovy "heads" (5 idioms), языки - jazyki "tongues" (2 idioms) in the plural section.

⁵In limited cases, mostly in idioms originally in the singular form, a higher degree of idiomaticity rendered necessary a random word.

our dataset consists of 438 sentences, divided into three different conditions: **Original** (73 with singular form and 73 with plural form), **Manipulated** and **Control** sentences as in the following table 2⁶.

4. Model and Experiment

Model For our experiment, we used the **ruBERT base model** (12-layer, 768-hidden, 12-heads, 178M parameters) provided by the Sberbank group (<https://huggingface.co/ai-forever/ruBERT-base>). This model was built by taking BERT [27] as a basis. ruBERT-base has a Byte-Pair Encoding (BPE) tokenizer with a dictionary of 120 thousand tokens. It was trained with 30 GB of Russian text, which includes Wikipedia, news, part of the Taiga corpus [28], and some books⁷.

Experiment In order to assess the sensitivity of the model to number variation in the selected idioms, we used the **pseudo-log-likelihood (PLL)** score [31]. As shown in Salazar et al. [6], the PLL can be considered as a measure comparable to probability. The authors demonstrate that the PLL outperforms scores obtained with auto-regressive models in a series of tasks related to sentence acceptability. Furthermore, in the work of Pedinotti et al. [7] the same measure is used to assess the difference in plausibility between metaphorical sentences, literal sentences and nonsense sentences. The results show a quite good correlation with human judgments of semantic plausibility.

The PLL of a sentence W can be derived by iteratively masking individual tokens, one at a time, using the main function of the MLM (Masked-Language Modeling) model. For each masked token w , the probability is calculated based on all other words in the context, and the log-probabilities for all tokens are summed. This process is illustrated by the following Equation 1:

$$PLL(W) = \sum_{t=1}^{|W|} \log P(w_t | W_{\setminus t}) \quad (1)$$

5. Results

We calculated the significance of the difference between the different conditions using the Wilcoxon test for multiple comparisons with Bonferroni correction.

⁶Unlike the control and the manipulated sentences, the translation of the original Russian idioms refers to the correspondent English idioms given by a Russian-English phraseological dictionary [26] whenever possible.

⁷We selected this model over DeepPavlov's ruBERT [29] because it performed better in the RussianSuperGLUE evaluation [30].

Table 2

Extracht from the dataset: **O** stands for **Original**, **M** for **Manipulated** and **C** for **Control** sentence.

Sentence	Number	Type
мне было очень стыдно за свой язык без костей - mne bylo očen' stydno za svoj jazyk bez kostej "I was very ashamed of my loose tongue ", lit. "I was very ashamed of my tongue without bones"	Sg	O
мне было очень стыдно за свои языки без костей - mne bylo očen' stydno za svoj jazyki bez kostej "I was very ashamed of my tongues without bones"	Pl	M
мне было очень стыдно за свое тело без костей - mne bylo očen' stydno za svoe telo bez kostej "I was very ashamed of my body without bones"	Sg	C
но все это для отвода глаз - no vse èto dlja otvoda galz "but all this is for a distraction", lit. "but all this is for the withdrawal of the eyes "	Pl	O
но все это для отвода глаза - no vse èto dlja otvoda glaza "but all this is for the withdrawal of the eye "	Sg	M
но все это для отвода вод - no vse èto dlja otvoda vod "but all this is for the withdrawal of the waters "	Pl	C

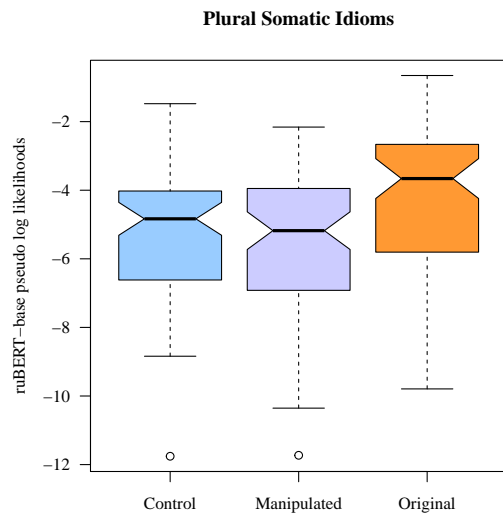


Figure 1: Boxplots of **ruBERT pseudo-log-likelihood** scores for the three conditions in the **plural** section.

Plural With regard to idioms containing somatic terms in the original plural condition, the differences between the **original condition** and the **manipulated condition**, and between the **original** and the **control condition** are significant ($p - value < 0.005$). On the other hand, there is no significant difference between the use of the singular (**manipulated condition**) and the different word in the **control condition** (we can also observe the differences in the Figure 1). This may be evidence of a real distance between the plural and singular forms.

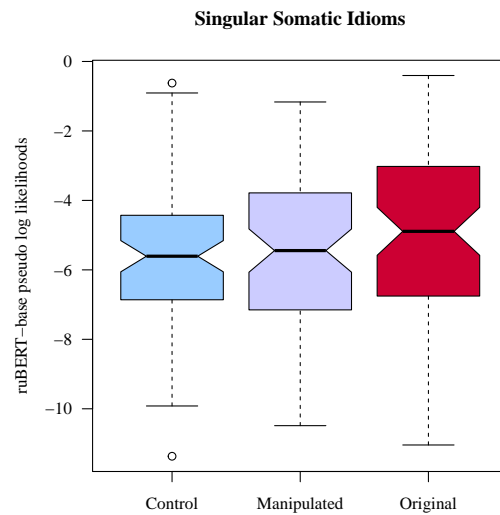


Figure 2: Boxplots of **ruBERT pseudo-log-likelihood** scores for the three conditions in the **singular** section.

As a matter of fact, the singular form proves to have a probability that is not significantly different from any other word occurring in the construction.

Singular As for the singular, we can observe a difference between the **original condition** and the other two (see Figure 2), but there is no significance from the statistical analysis we subjected the data to. The lack of significance could be due to a lower frequency of idioms featuring somatic parts in the singular form. Nonethe-

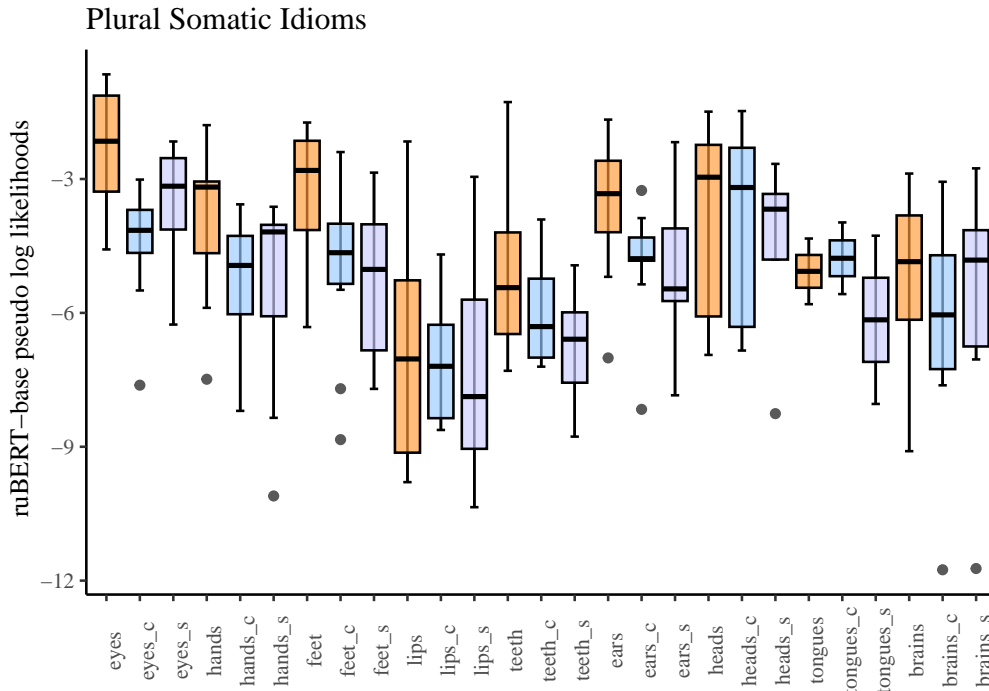


Figure 3: Boxplots of **ruBERT pseudo-log-likelihood** scores for the each noun in the three conditions in the **plural** section. The nouns are ordered by the relative frequency in the plural form as in Table 1.

less, the data show interesting trends when analysing the behaviour of individual nouns more in detail (as we can see in Figure 4).

6. Discussion

As it is evident from the results, we can positively address the first question, asserting that **the grammatical number of nouns significantly affects the structure of somatic idioms**. Indeed, the different probability obtained from the conditions indicate that the two grammatical values are not interchangeable as one would imagine. As a result, an overall morphological flexibility of the idiom associated with number variation as intended in [11] has not emerged.

The difference between the two values is significant in the **plural** section (see Figure 2), as it shows a higher probability to find a plural form in idioms originally in the plural (see Figure 1). On the other hand, the **singular** section displays a weak difference, except in two cases (“lip”, “tooth”), whose values result to be less interchangeable.

The differences between and within the two sections could be due to the morpho-semantic properties of the

grammatical number in the terms considered, specifically to their type of markedness. Firstly, **it could be plausible that the higher significance of the plural section correlates with local markedness**. As table 1 shows, most terms (six out of nine) display a condition of plural dominance, therefore the plural form is the naturally expected value. Secondly, in both sections these six locally marked terms show a difference between the **original condition** and the **manipulated condition**, demonstrating the impact of number. On the other hand, generally marked terms (“head”, “tongue”, “brain”) display little or no difference (see Figures 3 and 4).

Given the influence of the individual terms on both sections, it is worth discussing them separately.

Locally marked terms “Eye” and “ear” occupy a dedicated section in Ljaševskaja [14]’s investigation, as it was observed that they have similar contexts. This similarity seems to be confirmed both in the **plural** and in the **singular** section (see figure 3 and 4). However, at a closer look, “ear” behaves similarly to “hand/arm” and “foot/leg”, especially in the **plural** section. These nouns present a significant difference between the **original plural condition** and the **manipulated condition** (in

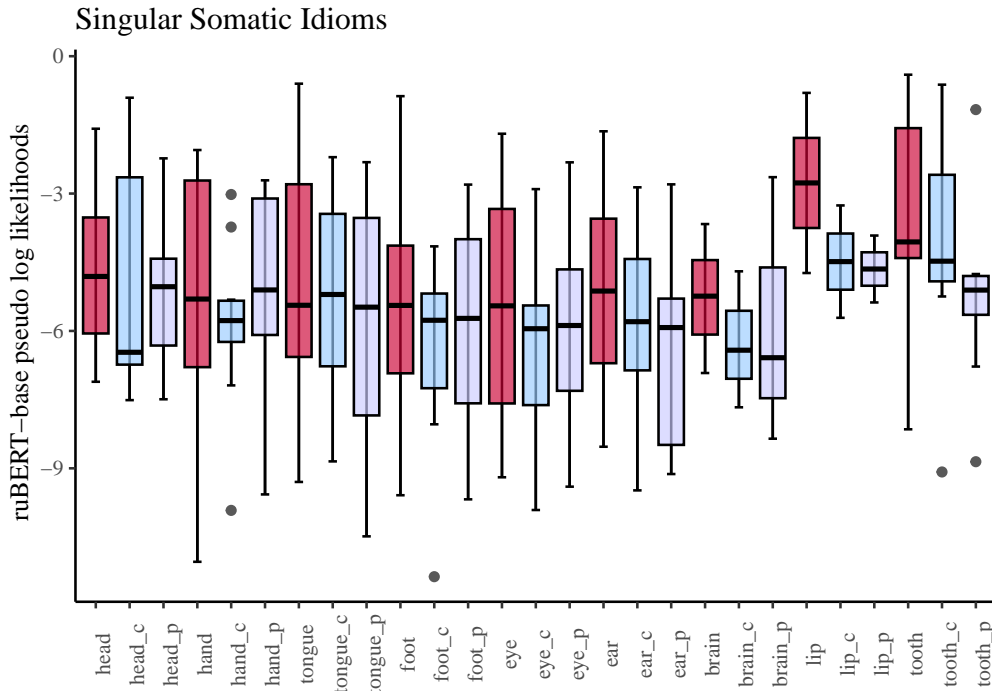


Figure 4: Boxplots of ruBERT pseudo-log-likelihood scores for the each noun in the three conditions in the **singular** section. The nouns are ordered by the relative frequency in the singular form as in Table 1.

“ears” and “hands” the p -value < 0.005 ; while in “feet” the p -value < 0.02 .

An unexpected outcome is offered by “lip”, therefore it cannot be associated with the nouns denoting a bipartite item, despite its referential meaning. Moreover, it behaves similarly to a noun denoting a complex body part: “tooth” (which is significant in both sections, p -value < 0.005). In the **singular** section “lip” is more probable in the singular value and it could be due to a more fixed structure of the original idiom.

Generally marked terms Among the generally marked terms “head” and “tongue” seem to follow the tendency observed in [11]. In fact, a certain interchangeability between the values is demonstrated by the **singular** section; while in the **plural** section, the original plural form is more probable, possibly denoting a more fixed structure.

Despite being a generally marked term, “brain”, provides an unexpected outcome. In fact, it features an evident difference between the values in the **singular** section, possibly alluding to a more fixed structure in the original idiom. On the other hand, little change is observed in the **plural** section. The results obtained in

the **singular** section do not explicitly confirm the existence of the additional sense associated with the plural of “brain” (as noted in Ljaševskaja [14, 86]). Nonetheless, the semantic discrepancy between the singular and the plural meaning of “brain” could still explain the less likely plural value in the **manipulated condition**.

7. Conclusions

On the basis of the presented experiment, we may assert that our initial assumptions prove to be partially justified. Comparing **original** sentences to **manipulated** and **control** sentences, we have found that, overall, number significantly affects the probability which connects the constituents of the idiom. Specifically, idiom variation appears to be subject to the morpho-semantic properties of the nominal constituent in the idioms considered. Indeed, our data showed that **the result varies depending on the type of markedness**. As a result, change in number for the generally marked constituents is more probable. This could confirm the general tendency of number being a flexibility dimension in idioms, which does not alter the idiomatic structure. On the other hand, locally marked terms show a significant difference in probabil-

ity between the two values. The difference suggests a less probable number variation, which may allude to a less flexible idiomatic structure. However, some terms exhibit a peculiar condition, whose interpretation may be subjected to the formulation of future hypotheses.

Limitations and future research Despite the correlation between human judgment and the PLL measure shown in Pedinotti et al. [7], the inclusion of human evaluations could still be necessary given the figurative meaning of idioms. This comparison could improve the robustness of the analysis while clarifying the results obtained by the model. Furthermore, given the unexpected outcome for some nouns, a larger dataset including more idioms per nouns and a wider variety of nouns could be designed.

References

- [1] I. Olza-Moreno, Aspectos de la semántica de las unidades fraseológicas. La fraseología somática metalingüística del español, Ph.D. thesis, Universidad de Navarra, 2009.
- [2] F. Vakk, O somatičeskoj frazeologiji v sovremennom èstonskom literaturnom jazyke, Ph.D. thesis, Institut jazyka i literatury Akademii nauk ÈSSR, 1964.
- [3] P. Acquaviva, Lexical Plurals: a morphosemantic approach, Oxford University Press, 2008.
- [4] P. M. Tiersma, Local and general markedness, *Language* (1982).
- [5] H. Baayen, C. Burani, R. Schreuder, Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian, *Yearbook of Morphology* (1997).
- [6] J. Salazar, D. Liang, T. Q. Nguyen, K. Kirchoff, Masked Language Model Scoring, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, p. 2699–2712. URL: <https://www.aclweb.org/anthology/2020.acl-main.240.pdf>.
- [7] P. Pedinotti, E. Di Palma, L. Cerini, A. Lenci, A howling success or a working sea? testing what BERT knows about metaphors, in: Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 192–204. URL: <https://aclanthology.org/2021.blackboxnlp-1.13>. doi:10.18653/v1/2021.blackboxnlp-1.13.
- [8] A. Cserép, Idiom variation and decomposability part i: Verbal variation, *Yearbook of Phraseology* (2017).
- [9] D. Dobrovol'skij, Windows to the Mind: Metaphor, Metonymy and Conceptual Blending, Walter de Gruyter, 2011.
- [10] A. Langlotz, Idiomatic Creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English, John Benjamins Publishing, 2006.
- [11] A. Cserép, Idiom variation and decomposability part ii: Variation in the noun phrase, *Yearbook of Phraseology* (2017).
- [12] V. Degtjarev, Kategorija čisla v slavjanskich jazykach. Istoriko-semantičeskoe issledovanie, Izdatel'stvo Južnogo Federal'nogo Universiteta, 2014.
- [13] V. V. Vinogradov, Russkij jazyk: grammatičeskoe učenie o slove, Vyšaja škola, 1986.
- [14] O. N. Ljaševskaja, Semantika russkogo čisla, *Jazyki Slavjanskoj Kul'tury*, 2004.
- [15] N. J. Švedova; Nina D. Arutjunova, et al., *Russkaja Grammatika*, Nauka, 1980.
- [16] D. Ušakov, Tolkovij slovar' russkogo jazyka, sov. encikl., 1935-1940.
- [17] D. G. T. Declerck, Towards the detection and formal representation of semantic shifts in inflectional morphology, in: 2nd Conference on Language, Data and Knowledge (LDK 2019), 2019.
- [18] A. Janszo, Disambiguating grammatical number and gender with bert, in: Proceedings of the Student Research Workshop Associated with RANLP 2021, 2021.
- [19] Y. Dai, Y. Liu, L. Yang, Y. Fu, An idiom reading comprehension model based on multi-granularity reasoning and paraphrase expansion, *Applied Sciences* 13 (2023). URL: <https://www.mdpi.com/2076-3417/13/9/5777>. doi:10.3390/app13095777.
- [20] V. Nedumpozhimana, F. Klubička, J. D. Kelleher, Shapley idioms: Analysing bert sentence embeddings for general idiom token identification, *Frontiers in Artificial Intelligence* 5 (2022). URL: <https://www.frontiersin.org/articles/10.3389/frai.2022.813967>. doi:10.3389/frai.2022.813967.
- [21] Z. Zeng, S. Bhat, Idiomatic expression identification using semantic compatibility, *Transactions of the Association for Computational Linguistics* 9 (2021) 1546–1562. URL: <https://aclanthology.org/2021.tacl-1.92>. doi:10.1162/tacl_a_00442.
- [22] J. Peng, A. Feldman, Automatic idiom recognition with word embeddings, in: Symposium on Information Management and Big Data, 2015.
- [23] A. N. B. D. O. Dobrovol'skij, *Akademičeskij slovar' russkoj frazeologii*, Leksrus, 2020.
- [24] A. I. Fedorov, *Frazeologičeskij slovar' russkogo literaturnogo jazyka*, Astrel', AST, 2008.
- [25] W. O'Grady, *The syntax of idioms*, *Natural Language Linguistic Theory* (1998).
- [26] S. Lubenskaja, Bol'soj russko-anglijskij frazeologičeskij slovar', AST-press, 2004.

- [27] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. arXiv:1810.04805.
- [28] S. O. Shavrina Tatiana, To the methodology of corpus construction for machine learning: «taiga» syntax tree corpus and parser., in: Proceeding of "CORPORA2017", international conference, 2017.
- [29] Y. Kuratov, M. Arkhipov, Adaptation of deep bidirectional multilingual transformers for russian language, 2019. arXiv:1905.07213.
- [30] T. Shavrina, A. Fenogenova, A. Emelyanov, D. Shevelev, E. Artemova, V. Malykh, V. Mikhailov, M. Tikhonova, A. Chertok, A. Evlampiev, Russian-superglue: A russian language understanding evaluation benchmark, arXiv preprint arXiv:2010.15925 (2020).
- [31] A. Wang, K. Cho, BERT has a Mouth, and It Must Speak: BERT as a Markov Random Field Language Model, in: Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation, 2019, p. 30–36. URL: <https://www.aclweb.org/anthology/W19-2304/>.