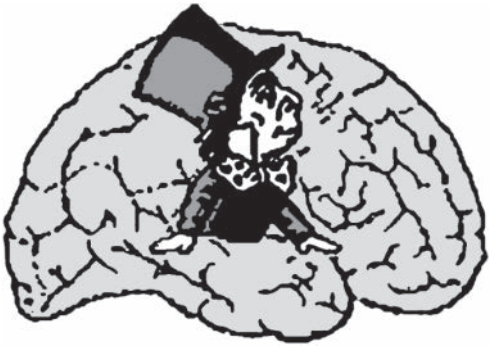
КОГНИТИВНАЯ НАУКА В МОСКВЕ

НОВЫЕ ИССЛЕДОВАНИЯ



МАТЕРИАЛЫ КОНФЕРЕНЦИИ 2017

ПОД РЕД. Е.В. ПЕЧЕНКОВОЙ, М.В. ФАЛИКМАН

УДК 159.9

ББК 81.002

К57

К57 Коллективный

Когнитивная наука в Москве: новые исследования. Материалы конференции 15 июня 2017 г.

Под ред. Е.В. Печенковой, М.В. Фаликман. – М.: ООО «Буки Веди», ИППиП. 2017 г. – 596 стр.

Электронная версия

ISBN 978-5-4465-1509-7

УДК 159.9

ББК 81.002

ISBN 978-5-4465-1509-7

DEMASKING RUSSIAN CASE INFLECTION

Vasilyeva M. D.
linellea@yandex.ru
Lomonsov MSU

Abstract. This study investigates the role of case in nominal wordform processing. Prior studies on Serbian, Finnish and German (e.g., Feldman, Fowler 1987; Laine et al., 1999; Clahsen et al., 2001) assume that the citation form is processed faster than other inflectional variants, but they are inconclusive with respect to whether oblique case forms behave alike or not. To this aim, we conducted a progressive demasking experiment with Russian singular nouns. We claim that the division of case forms into nominative and oblique is not sufficient and discuss possible reasons for the observed results.

Keywords: visual word recognition, morphological processing, progressive demasking, case syncretism, Russian declension

Studies on isolated visual word recognition suggest that the internal morphemic structure of a word affects its processing time course. For instance, oblique case forms are recognized more slowly than the nominative, the citation form (cf., Feldman, Fowler 1987 for Serbian; Laine et al., 1999 for Finnish). However, the exact nature of this effect is disputed. Early Serbian studies (e.g., Feldman, Fowler 1987) advocated for fundamental storage differences: in the mental lexicon, the nominative form functions as the nucleus of the inflectional paradigm, while other case forms, being satellites, surround it and, thus, are harder to be activated. The first Finnish findings were analyzed within the SAID model which attributes oblique case processing cost to the decomposition procedure, which is obligatory for inflection, but not for derivation (e. g., Laine et al., 1999). In later Finnish studies (e.g., Bertram et al., 2000) this effect was not replicated for wordforms embedded in sentences; the slow processing of case forms presented in isolation was therefore related to the absence or presence of appropriate context. Under all these approaches, no differences in oblique case form processing are postulated. Yet another viewpoint implies that syntactic functions fulfilled by the oblique wordform influence its processing (Serbian data by Kostić, 1991), and case syncretism plays a role as well (German data by Clahsen et al., 2001). The absence of a uniform account for the mechanism of case processing emphasizes the necessity of cross-language comparisons and favors experimental studies on languages that have not been previously examined. Russian is among such under researched languages.

In the present study, we tested whether case matters in Russian nominal singular wordform visual recognition. For our purposes we chose the progressive demasking task (PDT) instead of the widely employed lexical decision task (LDT). In PDT, a mask and the stimulus are presented consecutively in cycles, so that the duration of the mask decreases with the increase of stimulus duration. The stimulus gradually becomes more and more perceptually salient. The participant has to press a response button when he has identified the word. PDT is believed to tap into early stages of visual word recognition, while LDT is influenced by postlexical processes (Laine et al., 1999). Moreover, PDT does not require the usage of pseudowords. We focused only on inanimate nouns belonging to the two most productive Russian declensional classes: feminine *-a* and masculine *-ø* nouns (see Table 1) and compared response latencies to different singular case forms. We hypothesized that if the intrinsic properties of oblique case markers do not matter, we will obtain a difference only between nominative and other case forms for the two groups of nouns. If other factors such as ambiguity also play a role, we will observe differences between different oblique cases. As each noun group has its own pattern of inflection, oblique case processing may differ across inflectional class. We did not compare overt and zero inflection in the nominative case (feminine *-a* vs. masculine *-ø*), as this issue was already covered by (Gor et al., 2017).

Table 1. Endings for singular declension of the two most productive Russian inflectional classes.

	Nominative	Genitive	Dative	Accusative	Instrumental	Locative
feminine	-a	-y	-e	-u	-oj	-e
masculine	-ø	-a	-u	-ø	-om	-e

Method

48 right-handed native speakers of Russian took part in the study. We used 54 feminine and 54 masculine nouns matched for lemma frequency. All stimuli were base nouns; they did not undergo any stem alternations and had fixed stress on the stem. Length in the nominative form differed from 4 to 6 letters (each group comprised one third of words with each length). A Latin square design was employed with the number of lists corresponding to the number of case forms[1].

Each participant was assigned to one of the six experimental lists and was tested individually. Stimuli were presented using PsycoPy software (Peirce, 2009). In each trial, the mask (a row of hash marks) and the target stimulus were presented sequentially in cycles. The duration of the cycle was held constant (210 ms). In the first cycle the duration of the mask was 195 ms, and the duration of the

[1] We decided against prior presentation of a disambiguating context, as context is believed to alleviate the decomposition processing cost (Laine et al., 1999). As our words were presented without context, case labels for ambiguous endings (*-y* and *-e* for feminine nouns, *-ø* for masculine nouns) are somewhat arbitrary. Hence, we do not expect any differences between locative and dative *-e*, nor nominative and accusative *-ø*. However, this is needed for counterbalancing issues, as patterns of syncretism do not coincide across our two noun groups.

stimulus was 15 ms. With each successive cycle, the duration of the mask decreased by 15 ms, and the duration of the stimulus increased by 15 ms. The cycles continued until the participant hit the spacebar indicating that s/he has recognized the word. After that, s/he had to type the word so that we could check the accuracy of identification.

Results

Accuracy data was not analyzed (incorrect responses constituted less than 5% of the data). Prior to the RT analysis, incorrect and too slow (> 3000 ms) responses were removed. We applied log-transformation to reduce the positive skew. Remaining outliers were cut off via interquartile trimming by participants, items (lexemes), gender and case.

We used mixed effects modeling for the analysis of reaction times as implemented in the package lme4 in the statistical software R. *T* values, *p* values and standard errors were determined using the package lmerTest. Fixed and random effects were included only if they significantly improved the model's fit in a backward stepwise model selection procedure. Models were selected using chi-square log-likelihood ratio tests with regular maximum likelihood parameter estimation. Subject and lexeme were treated as random effects. Trial order (*z* transformation on log numbers) was included to control for longitudinal task effects such as fatigue or habituation; experimental list was included to avoid potential counterbalancing issues. Lexeme and wordform frequency[2], length in letters and syllables, mean Levenstein distance to the nearest 20 lexeme-neighbors[3], inflectional and relative entropy measures[4] (Milin et al., 2009) were additionally included as covariates. All these counts were log-transformed. To avoid multicollinearity, all counts except for trial were transformed into five principle components, explaining 93.5% of the variance. The first principle component (PC1) captured orthographic characteristics of the stimulus. The second component (PC2) was inversely related to frequency. The third component (PC3) was inversely related to relative entropy and positively related to inflectional entropy. Paired contrasts were carried out in the package lsmeans. For paired comparisons, FDR adjusted p-values are reported.

The final model included the following fixed factors: trial ($\chi^2(1) = 71.47$, $p < .001$), PC2 ($\chi^2(1) = 8.11$, $p = .004$), case ($\chi^2(10) = 142.63$, $p < .001$), gender ($\chi^2(6) = 32.26$, $p < .001$) and case by gender interaction ($\chi^2(5) = 29.99$, $p < .001$); see Fig. 1. All other predictors and interactions turned out to be insignificant.

Feminine nouns. Nominative was recognized faster than oblique cases: dative ($t(4811.3) = -6.89$, $p < .001$), accusative ($t(4870.7) = -2.96$, $p = .005$), instrumental ($t(4788.8) = -2.63$, $p = .014$) and locative ($t(4811.7) = -5.98$, $p < .001$); the difference

[2] Lexeme frequency was taken from the frequency dictionary by Lyashevskaya and Sharoff (2009). Worform frequency was manually extracted from the Russian national corpora (www.ruscorpora.ru). To avoid zero frequencies, one was added to all counts.

[3] It was calculated in the vwr package (Keuleers, 2013).

[4] Exponent frequency was taken from the database by Slioussar and Samojlova (2014).

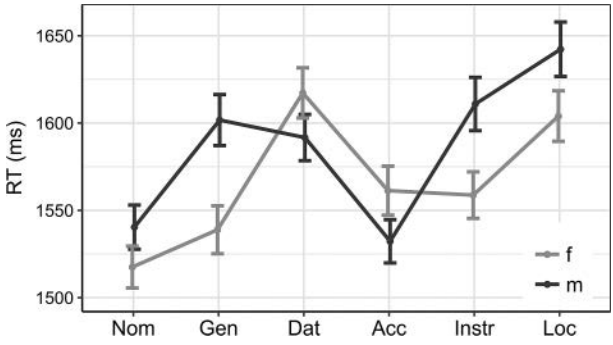


Figure 1. Raw RTs (with standard error bars) as a function of gender and case: *f* stands for feminine nouns, *m* stands for masculine nouns

from genitive was marginally significant ($t(4834.9) = -1.94$, $p = .073$). RTs to genitive wordforms did not differ significantly from accusative ($t(4818.9) = -0.99$, $p = .383$) and instrumental ($t(4879.2) = -0.63$, $p = .574$) forms, but were faster than RTs to dative ($t(4875.5) = -4.85$, $p < .001$) and locative ($t(4875.4) = -3.96$, $p < .001$) forms. Instrumental was recognized faster than dative ($t(4786.2) = -4.25$, $p < .001$) and locative ($t(4786.3) = -3.35$, $p = .002$); the difference from accusative was marginally significant ($t(4838.6) = -0.35$, $p = .739$). Accusative was recognized faster than dative ($t(4808.8) = -3.89$, $p < .001$) and locative ($t(4808.7) = -3$, $p = .005$) forms. The difference between dative and locative forms was not significant ($t(4779.2) = 0.88$, $p = .439$).

Masculine nouns. Nominative and accusative RTs did not differ significantly ($t(4777.4) = 0.41$, $p = .709$), but were smaller than RTs to other case forms: genitive ($t(4821.4) = -4.21$, $p < .001$ and $t(4823.3) = -4.64$, $p < .001$), dative ($t(4833.8) = -3.73$, $p < .001$ and $t(4831.9) = -4.16$, $p < .001$), instrumental ($t(4778.4) = -4.56$, $p < .001$ and $t(4778.8) = -4.99$, $p < .001$) and locative ($t(4784.1) = -6.7$, $p < .001$ and $t(4782.9) = -7.14$, $p < .001$). RTs to dative wordforms did not differ significantly from RTs to instrumental ($t(4839.3) = -0.8$, $p = .481$) and genitive ($t(4880.1) = -0.47$, $p = .682$) forms, but were faster than locative ($t(4807.6) = -2.96$, $p = .005$) forms. Locative was recognized slower than instrumental ($t(4787.8) = 2.17$, $p = .043$) and genitive ($t(4847.9) = 2.49$, $p = .02$) forms. Genitive and instrumental wordform RTs did not differ significantly ($t(4816.8) = -0.33$, $p = .739$).

Discussion

Results of our experiment replicate the effect attested for other languages: the nominative case is recognized faster than other case forms. Additionally, we observed differences between oblique case forms. Masculine and feminine wordforms ending in *-e* require much more time to be identified than other oblique case forms. Feminine *-e* is ambiguous between dative and locative cases. But ambiguity alone cannot be a source for this effect, as the recognition of another feminine wordform, *-y*, which is also ambiguous (genitive singular / nominative plural),

is not impaired. Actually, their RTs barely differ from the nominative singular, which implies that the nominative plural is the default interpretation. In theoretical morphology, it has been claimed that the ambiguity of *-e* is not accidental and that these forms have one underspecified representation (i.e., Müller, 2004). However, prior research on German adjectival inflection (i.e., Clahsen et al., 2001) suggests that underspecified forms are processed faster. Masculine *-e* wordforms (locative), on the contrary, are not ambiguous, but Russian locative is always governed by a preposition and in the present study locatives were presented without prepositions in the experimental conditions. We argue that our findings could be accounted for in a model of Russian case where all *-e* forms have one shared superspecified representation. The features distinguishing between two cases (dative vs. locative) compete with each other during wordform processing, and this competition slows down word identification in an environment without context. Thus, our results disprove models residing solely on the absence or presence of context (Bertram et al., 2000) or affix stripping processing cost (Laine et al., 1999) or the equal storage status of obliques (Feldman, Fowler, 1987), and extend the outreach of theories relying on morphosyntactic features (e. g., Clahsen et al., 2001) to the Russian language.

References

Bertram R., Hyönä J., Laine M. The role of context in morphological processing: Evidence from Finnish // Language and Cognitive Processes. 2000. Vol. 15. No. 4/5. P. 367 – 388. doi:10.1080/01690960050119634

Brysbaert M., Diependaele K. Dealing with zero word frequencies: A review of the existing rules of thumb and a suggestion for an evidence-based choice // Behavior Research Methods. 2013. Vol. 45. No. 2. P. 422 – 430. doi:10.3758/s13428-012-0270-5

Clahsen H., Eisenbeiss S., Hadler M. Sonnenstuhl I. The mental representation of inflected words: An experimental study of adjectives and verbs in German // Language. 2001. Vol. 77. P. 510 – 543.

Feldman L. B., Fowler C. A. The inflected noun system in Serbo-croatian: Lexical representation of morphological structure // Memory & Cognition. 1987. Vol. 15. P. 1 – 12.

Gor K., Chrabaszcz A., Cook S. Processing of native and nonnative inflected words: Beyond affix stripping // Journal of Memory and Language. 2017. Vol. 93. P. 315 – 332. doi:10.1016/j.jml.2016.06.014

Keuleers E. Vwr: Useful functions for visual word recognition research. R package version 0.3.0.. URL: http://CRAN.R-project.org/package.

Kostić A. Informational approach to the processing of inflected morphology: Standard data reconsidered // Psychological Research. 1991. Vol. 53. P. 62 – 70.

Laine M., Vainio S., Hyönä J. Lexical Access Routes to Nouns in a Morphologically Rich Language // Journal of Memory and Language. 1999. Vol. 40. P. 109 – 135.

Lyashevskaya O. N., Sharoff S. A. New dictionary of Russian frequencies [In Russian, 'Novyj chastotnyj slovar' russkoj leksiki']. M.: Azbukovnik, 2009.

Milin P., Đurđević D. F., Moscoso del Prado M. F. The simultaneous effects of inflectional paradigms and classes on lexical recognition: evidence from Serbian // Journal of Memory and Language. 2009. Vol. 60. No. 1. P. 50 – 64.

Müller G. On Decomposing inflection class features: syncretism in Russian noun inflection // Explorations in nominal inflection / G. Müller, L. Gunkel, G. Zifonum (Eds.). Berlin: Mouton De Gruyter, 2004. P. 189 – 228.

Peirce J. W. Generating stimuli for neuroscience using PsychoPy // Frontiers in Neuroinformatics. 2009. Vol. 2. P. 1 – 8. URL: http://journal.frontiersin.org/article/10.3389/neuro.11.010.2008/full.

Slioussar N., Samojlova M. Frequencies of different grammatical features and inflectional affixes of Russian nouns: a database 2014. URL: http://www.slioussar.ru/freqdatabase.html..

Русское падежное словоизменение: снимая маски

М. Д. Васильева
linelleamv@gmail.com
МГУ имени М.В. Ломоносова, Москва

Аннотация. В данной работе исследуется роль падежа в обработке словоформ существительных. Предыдущие исследования на материале сербского, финского и немецкого языков (см., в частности, Feldman, Fowler, 1987; Laine et al., 1999; Clahsen et al., 2001) показывают, что номинатив обрабатывается быстрее других падежных форм, сведения же о наличии различий в обработке форм косвенных падежей противоречивы. Поэтому мы провели эксперимент с русскими существительными в единственном числе, используя метод постепенной демаскировки. Согласно нашим данным, деления падежных форм на номинатив и косвенные падежи недостаточно. Мы обсуждаем возможные варианты анализа полученных результатов.

Ключевые слова: зрительное восприятие слов, морфологическая обработка, постепенная демаскировка, падежный синкретизм, русское склонение