

## Gender variation and markedness in Russian.

In this paper, we offer a new theoretical approach to gender markedness hierarchy in modern Russian. We base our analysis on corpus and experimental data on Russian nouns that demonstrate high degrees of variation in gender. We analyze factors that correlate with gender assignment the most, namely: declension class, lexical gender (if any), phonological form, and surface form analogy. We conclude that neuter is the unmarked gender language-wide, while masculine demonstrates some default-like behavior due to its frequency and diversity of surface forms. We call it a *greedy class* using an analogy to greedy operators in regular expressions in programming.

**General distribution.** Masculine is the most frequent gender in Russian (49% of all nouns, see Table 1). It is also the most diverse class in terms of surface forms: there are masculine nouns ending in nominative in a hard consonant, soft consonant, *-a*, *-o*, and *-e* (expressive forms, e.g., *domiško* ‘little house’). There are masculine indeclinable nouns ending in other vowels (e.g., *viski* ‘whiskey’, *kenguru* ‘kangaroo’). Neuter nouns are the smallest class (18%) and end in *-o* or *-e*. Indeclinable neuter nouns may end in other vowels (e.g. *džiu-džitsu* ‘jiu-jitsu’) Neuter agreement appears in impersonal sentences (e.g., *SvetaloN*. ‘It was getting bright’) which indicates the absence of the gender feature. Feminine nouns are 29% of all nouns and end in *-a* or a soft consonant. One can also see in Table 1 that gender and declension classes are tightly related in Russian.

**Expressive nouns.** We made corpus and experimental study of gender assignment to nouns with ambiguous or contradictory gender cues. In expressive nouns, the conflicting cues are the lexical gender of the base noun and the final segment of the expressive suffix that defines the declension class of the derived noun. For example, noun *vinčik* ‘wineDIM’ has neuter lexical gender and ends in a hard consonant, which is associated with masculine. We used the web-as-a-corpus and then conducted an experimental study of these nouns. Stimuli were balanced by the base gender, the inflection of the derivate, and its animacy. We collected 1200 answers. In 614 out of them (51%), expressive derivates did not preserve the base gender. The feminine base gender was lost in 461 cases (77%), while the masculine base gender was lost only in 153 cases (26%). Base gender changes were predominantly triggered by the inflectional affix, as expected (92% of cases). Interestingly, in 47 answers, masculine gender was assigned to nouns with feminine base gender and suffixes ending in *-o* or *-e* (class IV, associated with neuter). Masculine gender has a special status: it is preserved significantly more often and is triggered by changing the inflection class significantly more often than feminine or neuter. Moreover, it can be assigned in case of conflicting cues that do not suggest masculine.

**Indeclinable nouns.** According to a dictionary study by Murphy (2000), 67% of indeclinable nouns are listed as neuter, 17% as masculine, 8% as feminine, and the remaining nouns show variation. We analyzed data from a LiveJournal part of Russian General Internet Corpus. For this study, we had 137 nouns and 34214 observations in the final dataset. Indeclinable nouns are traditionally viewed as a class preferring neuter (Corbett & Fraser 1993) - neuter is indeed the most frequent gender (43%), but it is closely followed by masculine (37%). Feminine is 20% of instances. This is different from (Murphy 2000), although the hierarchy of genders is the same. The significant factors for gender assignment appeared to be the final segment (which here is a part of the root) and the gender of the semantic analogy of a given noun. 88 of the 137 nouns in our dataset have one salient semantic analogy, 52 can be associated with masculine, 21 with feminine, and 15 with neuter. Neuter is often assigned in the absence of any clear cues.

**Discussion.** These findings give us a basis for a theoretical analysis of gender markedness hierarchy. We propose that neuter is the unmarked gender in Russian language-wide (as in structural approaches e.g. Adamson & Šereikaitė 2019, Kramer 2015, Nevins 2011), while masculine is *greedy* as it contains an analogy to almost any surface form because it is the most frequent gender and the most diverse in terms of surface forms. Many patterns found in feminine and neuter can also be found in masculine: there are masculine nouns with all kinds of surface forms, while feminine and neuter are more restricted (see Table 1) At the same time, neuter is representationally unmarked and is chosen when there are no gender assignment cues. If cues are present, the ones that point to masculine will tend to win over the others. We call masculine a *greedy class* using an analogy to greedy operators in regular expressions in programming.

Table 1. A system of four inflectional classes for Russian nouns.

Classes	Descriptions	Examples	% in the RNC <sup>1</sup>
class I	M nouns ending in a C in Nom.Sg	<i>zakon</i> ‘law’, <i>kon</i> ‘horse’	46%
class II	F nouns ending in <i>-(j)a</i> in Nom.Sg	<i>komnata</i> ‘room’, <i>zemlja</i> ‘earth’	29%
	M nouns ending in <i>-(j)a</i> in Nom.Sg	<i>papa</i> ‘dad’, <i>djadja</i> ‘uncle’	1%
class III	F nouns ending in a C in Nom.Sg	<i>kost</i> ‘bone’	5%
class IV	N nouns ending in <i>-o</i> or <i>-e</i> in Nom.Sg	<i>okno</i> ‘window’, <i>more</i> ‘sea’	18%
indeclinable	nouns of different genders	<i>kivi</i> ‘kiwi’, <i>pal’to</i> ‘coat’	1%

Adamson, Luke and Milena Šereikaitė. 2019. Gender Representation and Defaults in Lithuanian. *Glossa: a journal of general linguistics* 4(1): 121. 1–35.

Corbett, Fraser 1993 — Corbett, G. G., Fraser, N. M. Network Morphology: a DATR account of Russian nominal inflection. *Journal of linguistics*, 1993, 29: 113-142.

Murphy, Dianna L. (2000). “The Gender of Inanimate Indeclinable Common Nouns in Modern Russian”. University of Ohio: Doctoral dissertation.

Kramer 2015 — Kramer, R. T. The morphosyntax of gender. Oxford: Oxford University Press, 2015.

Nevins, Andrew. 2011. "Marked targets versus marked triggers and impoverishment of the dual." *Linguistic Inquiry* 42.3 : 413-444.

---

<sup>1</sup> Percentages of nouns in the Russian National Corpus, or RNC