Gender and case in Russian nouns denoting professions and social roles

Varvara Magomedova
varya.magomedova@gmail.com
independent researcher

Natalia Slioussar
slioussar@gmail.com
HSE, Moscow
SPbU, Saint Petersburg

Outline

- Nouns like *vrač* ‘doctor’: historically M, but now allow for M and F agreement (e.g. *etot / eta vrač*).
- We focus on a complex relationship between **gender and case** features.
- **Previous studies:** F is grammatical only in Nom, although other forms are attested.
  (e.g. Graudina et al. 1976; Zaliznjak 2002; Sitchinava 2011)
- **Our study:** how often is F agreement is produced and how easily is it processed depending on case? Web-as-a corpus approach and three experiments.

Previous studies

- No corpus studies, experimental studies focus on agreement with Nom forms.
- Panov (1968) and Novikov & Priestly (1999): the choice of M/F in agreeing **verbs and adjectives**. Corbett (2006) incorporated this in a more general theory of agreement.
- **Language acquisition studies** (Dizer 2007; Dobrova 2013; Rodina & Westergaard 2012; Rodina 2014; Tseitlin 2009): children acquire semantic agreement relatively late.

Previous studies

- Garnham and Yakovlev (2015): a list of 160 nouns (stereotypes, paired/unpaired). A processing experiment (sentence-by-sentence reading times).
- Several studies by Slioussar et al. (Slioussar & Generalova 2018 etc.): various processing experiments (grammaticality judgment and word-by-word reading times).

Web-as-a corpus study

- 43 unpaired nouns from G&Y list: 42 ending in a consonant like *psixolog* ‘psychologist’ + *sud’ja* ‘judge’.
- **Method:** searching for M/F pronouns (*moj* ‘my’, *naš* ‘our’, *etot* ‘this’) + nouns in all six cases in singular.
- Google search engine, very approximate results (mostly checking what is attested at all, rather than frequencies).
- **Results.** 30 nouns: less than 5000 hits with F agreement, less than 30 oblique forms. Some stereotypically M professions like *mexanik* ‘mechanic’: **no** F agreement.

  13 nouns selected for further analysis.

Web-as-a corpus study

- Two examples with F and M agreement:

M agreement	Nom	Gen+Acc	Dat	Ins	Loc	Total
fotograf 'photographer'	335400 (48.2%)	163720 (23.5%)	101600 (14.6%)	91540 (13.2%)	3317 (0.5%)	695577
sudja 'judge'	102300 (28.3%)	188669 (52.2%)	17250 (4.8%)	38700 (10.7%)	14548 (4.0%)	361467

F agreement	Nom	Gen	Dat	Acc	Ins	Loc	Total
fotograf 'photographer'	44600 (99.1%)	265 (0.6%)	62 (0.1%)	43 (0.1%)	43 (0.1%)	0	45013
sudja 'judge'	14430 (40.0%)	7850 (21.8%)	5609 (15.6%)	5396 (15.0%)	2169 (6.0%)	574 (1.6%)	36028

- Conclusion. The share of oblique forms: F<<M. The only exception is *sudja*. Percentages vary, but F forms in Loc are especially infrequent (often unattested).

Experiment 1: grammaticality judgement

- **Participants**: 53 native Russian speakers.
- **Materials**: 15 unpaired nouns denoting stereotypically feminine professions, 5 sentences with each noun, 5 oblique cases, 5 experimental lists (15 targets + 30 fillers).

(1) a. *Ja uznal o svoem diagnoze ot našej vrača.*
I learned about self's diagnosis from our$_{F.GEN.SG}$ doctor$_{GEN.SG}$
'I learned about my diagnosis from our doctor'.

b. *Ja obratilsja s etoj problemoj k našej vraču.*
I appealed with this problem to our$_{F.DAT.SG}$ doctor$_{DAT.SG}$
'I asked our doctor about this problem'.

- **Method**: judging sentence grammaticality on a 1 to 5 scale. The IbexFarm platform (www.spellout.net).

Experiment 1: grammaticality judgement

- Results: average ratings: 2.0 for Gen, 2.0 for Dat,1.9 for Acc, 2.0 for Ins, 1.8 for Loc.
- Statistics here and below: ordinal logistic or linear regressions with mixed effects (intercepts) by participant and by item. No significant differences in this experiment.
- Conclusion: all oblique forms were judged as equally marginal. This agrees with our corpus study and with previous research. But a more sensitive method may zoom on the differences between these forms.

Experiment 2: ranging sentences

- Participants: 35 native Russian speakers.
- Materials: 6 nouns in 30 sentences from Exp.1.
- Method: 5 sentences with one noun in different cases are presented together (in a random order). Participants are asked to range them from the best to the worst using a 1 to 5 scale. The *PsychoPy* software (www.psychopy.org).
- **Results:** average ratings: 4.0 for Ins, 3.4 for Acc, 3.0 for Gen, 2.9 for Dat and 1.4 for Loc.
- Conclusion: Loc is significantly worse than other oblique cases.

Loc vs. Acc: β=-4.38, SE=0.41, z=-10.69, p<0.01; Loc vs. Dat: β =-3.34, SE=0.30, z=-11.16, p<0.01;
Loc vs. Gen: β=-4.05, SE=0.37, z=-11.06, p<0.01; Loc vs. Ins: β=-3.71, SE=0.33, z=-11.22, p<0.01.

Experiment 3: self-paced reading

- **Participants**: 68 native Russian speakers
- **Materials**: 24 unpaired nouns, 48 sentences in two conditions (M / F agreement), 6 cases, 2 experimental lists (48 targets + 108 fillers).

(2) a. *Za stolom sidit molodoj / molodaja bibliotekar' v sinem pidžake.*
at table sits young$_{M.NOM.SG/F.NOM.SG}$ librarian$_{NOM.SG}$ in blue jacket.
'A pretty librarian wearing a blue jacket is sitting at the table'.

b. *Petr uznal ot opytnogo / opytnoj vrača o svoem diagnoze.*
Peter learned from experienced$_{M.GEN.SG/F.GEN.SG}$ doctor$_{GEN.SG}$ about self's diagnosis.
'Peter learned about his diagnosis from an experienced doctor'.

c. *Vanja priglasil populjarnogo / populjarnuju dietologa na večernee šou.*
Vanya invited popular$_{M.ACC.SG/F.ACC.SG}$ dietologist to evening show
'Vanya invited a popular dietologist to the evening show.'

Experiment 3: self-paced reading

- **Method**: self-paced reading with comprehension questions after 1/3 sentences. The IbexFarm platform (www.spellout.net).

  --- ----. You ----. --- read.

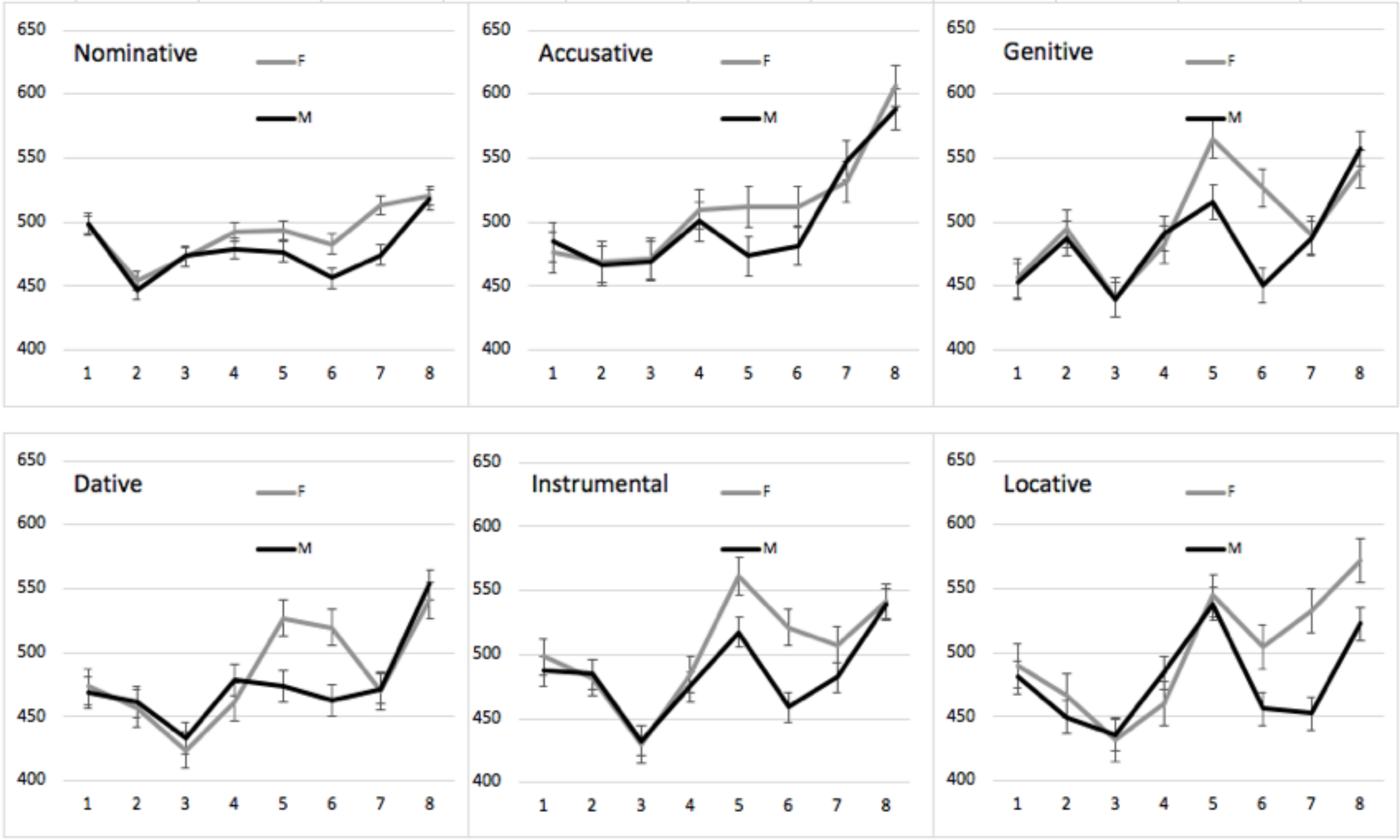
- **Results.** Target noun (N region): significant differences in the Gen, Dat and Ins groups. N+1 region: significant differences in every group. N+2 and N+3 regions: significant differences only in the Loc group.

- Conclusion: Loc is the most problematic.

N region: β=40.08, SE=19.04, z=2.02, p=0.04 for Gen, β=39.15, SE=17.91, z=2.19, p=0.03 for Dat, β=43.33, SE=20.78, z=2.09, p=0.04 for Ins. N+1 region: β=26.43, SE=12.55, z=2.11, p=0.04 for Nom; β=48.02, SE=12.33, z=3.90, p<0.01 for Gen; β=66.22, SE=14.84, z=4.46, p<0.01 for Dat; β=37.01, SE=15.24, z=2.43, p=0.02 for Acc; β=61.01, SE=13.80, z=4.42, p<0.01 for Ins; β=37.80, SE=14.57, z=2.59, p=0.01 for Loc. N+2 region: β=67.79, SE=11.87, z=5.71, p<0.01 for Loc. N+3 region: β=49.04, SE=17.82, z=2.75, p<0.01 for Loc.

Experiment 3: self-paced reading

Figures 1-6. Average word-by-word reading times (in ms).



Discussion: form matters!

- Why is F agreement marginal in oblique cases? A deep connection between gender and declension.

  Hard to explain in many morphological theories (e.g. Kramer 2015). Some models have better chances (e.g. Rice 2005; Doleschal 2000), but did not address this.

- Why is Loc especially bad? Most probably, syncretism.

  Interesting both for theoretical morphology and for psycholinguistics. Previous studies: facilitatory effects of syncretism!

  Kak budto eto ne vrač, a kakaja-to vrača!
  as if this not doctor$_{\text{NOM.SG (1D)}}$ but some$_{\text{F.NOM.SG}}$ doctor$_{\text{NOM.SG (non-existent 2D noun)}}$

Discussion: form matters!

- Some examples from Slioussar (2018) finding facilitatory effects of syncretism:

(3) a. *Bilet na koncert / koncerty byli…*
ticket for concert$_{ACC.SG}$ / $_{ACC.PL=NOM.PL}$ were

b. *Komnata dlja večerinki / večerinok byli…*
room for party$_{GEN.SG=NOM.PL}$ / $_{GEN.PL}$ were

c. *Logotip na futbolke / futbolkax byli…*
logo on T-shirt$_{LOC.SG}$ / $_{LOC.PL}$ were

- Form matters! See also other work by Slioussar and Magomedova (on gender agreement errors with different declensions, on expressive nouns), a recent corpus study by Chuprinko and Kholodilova (on undeclinable nouns).