

Babby's puzzle: syntax or morphology?



Asya Pereltsvaig
(Stanford)
FASL 2009, Cornell

“The area of Russian quantified phrases is characterized by a very *significant variation* in norm and usage; perhaps this is the most unstable area of modern Russian as far as grammatical norm is concerned.”

Mel'čuk (1985: 106), translation and emphasis mine



The goal of this talk: to disentangle and to explain some of this variation and complexity

Babby's puzzle: heterogeneous vs. homogeneous case distributions (cf. Babby 1987)

- Heterogeneous (in direct case positions)

(1) Vanya napisal pjat' knig.
Vanya wrote five.**ACC** books.**GEN**
'Vanya wrote five books.'

- Homogeneous (in oblique case positions)

(2) Vanya gorditsja pjat'ju knigami.
Vanya is-proud-of five.**INSTR** books.**INSTR**
'Vanya is proud of five books.'



In this talk:

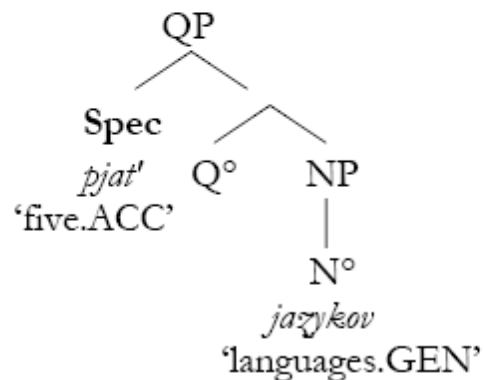
- Syntactic analysis of the heterogeneous/homogeneous contrast (Bailyn 2004, Pereltsvaig 2006b)
- Problematic facts concerning paucal numerals
 - large-scale study of current usage
- Proposed account

Syntactic approaches to Babby's puzzle

- Bailyn (2004), Pereltsvaig (2006b)

Heterogeneous case distribution

- obtains in direct case positions (NOM and ACC)
- Numerals as **Spec-QP**:



Homogeneous case distribution

- obtains in oblique case positions (DAT, INSTR, LOC, GEN)
- Numerals as **Q°**:



Arguments for the dual analysis (1)

- from Bailyn (2004):
 - Phrasal (idiomatic) quantity expressions can appear only with the heterogeneous case distribution
 - Under the dual analysis: with the homogeneous case distribution a PP cannot be merged in Q°

(3) a. Džejms Bond znaet [PP do figa] jazykov.
James Bond knows to fig languages.**GEN**
'James Bond knows a lot of languages.'

b. *Džejms Bond vladeet [PP do figa] jazykami.
James Bond knows to fig languages.**INSTR**
'James Bond knows a lot of languages.'

Arguments for the dual analysis (2)

- from Pereltsvaig (2006):
 - Approximative Inversion (which is independently argued to be HM; cf. Pereltsvaig 2006b) is possible only with the heterogeneous case distribution
 - Under the dual analysis: with the homogeneous case distribution the numeral is a head Q° and hence it blocks Head Movement (= HMC violation)

- (4) a. Džejms Bond znaet jazykov pjat'.
James Bond knows languages.**GEN** five.**ACC**
'James Bond knows something like five languages.'
- b. *Džejms Bond vladeet jazykami pjatju.
James Bond knows languages.**INSTR** five.**INSTR**
'James Bond knows something like five languages.'



Interim conclusion

- The hetero-/homogeneous case distribution correlates with (or results from) the different merger of the numeral (Spec-QP vs. Q°).
- For Ionin & Matushansky (2004, 2006) in the heterogeneous pattern the numeral is merged as a head that assigns/checks GEN

In the heterogeneous pattern, why genitive?

- Baily (2004):
 - The Case checked in the heterogeneous pattern is just like other types of genitives in Russian (e.g., Gen of Negation, Partitive Gen, Intensional Gen, Adnominal Gen, Comparative Gen, etc.)
 - Overextending the analysis?
- Ionin & Matushansky (2004, 2006):
 - Numerals are Ns; hence, The Case checked in the heterogeneous pattern is just like Adnominal Gen
 - But cf. Pereltsvaig (2007) for arguments against treating numerals as Ns

Further complications: paucal numerals

- Paucal numerals: 2, 3, 4, complex numerals ending in 2-4, *pol* '1/2' and *poltora* '1 1/2'
- cf. Babby (1987), Franks (1995, 1998), Ionin & Matushansky (2004, 2006), Rappaport (1998, 2004), a.o.

Paucal Numerals

- Paucal numerals occur (in the heterogeneous pattern) with nouns in GEN.SG (rather than GEN.PL, as with 5&Ups) – *or do they?*

(5) a. Vanya napisal pjat' knig. → 5&Ups

Vanya wrote five.ACC books.**GEN.PL**

'Vanya wrote five books.'

b. Vanya napisal tri knigi. → Paucal

Vanya wrote three.ACC books."**GEN.SG**"

'Vanya wrote three books.'

- Bailyn (2004), Ionin & Matushansky (2004, 2006): selectional property of paucal numerals

But is “GEN.SG” really genitive singular?

- There are good reasons to believe that this form is neither GEN, nor SG:
 - special count forms
 - agreement/anaphora
 - processing
 - non-GEN.SG forms of certain Ns and Adjs

Paucal count forms in other East Slavic languages

(8) Ukrainian (with masculine nouns):

dva {bráty / *bráta/*bratý / *bratív}

two brother.QUANT / *brother.GEN.SG
/*brother.NOM.PL /*brother.GEN.PL

'two brothers' (Shevelov 1993: 988; cf. also
Mel'čuk 1985: 436)

(9) Belorussian (with neuter nouns):

try {sjalý / *selá /*sëly/*sël}

three village.QUANT / *village.GEN.SG
/*village.NOM.PL /*village.GEN.PL

'three villages' (Mayo 1993: 935)

Agreement/Anaphora with paucal subjects

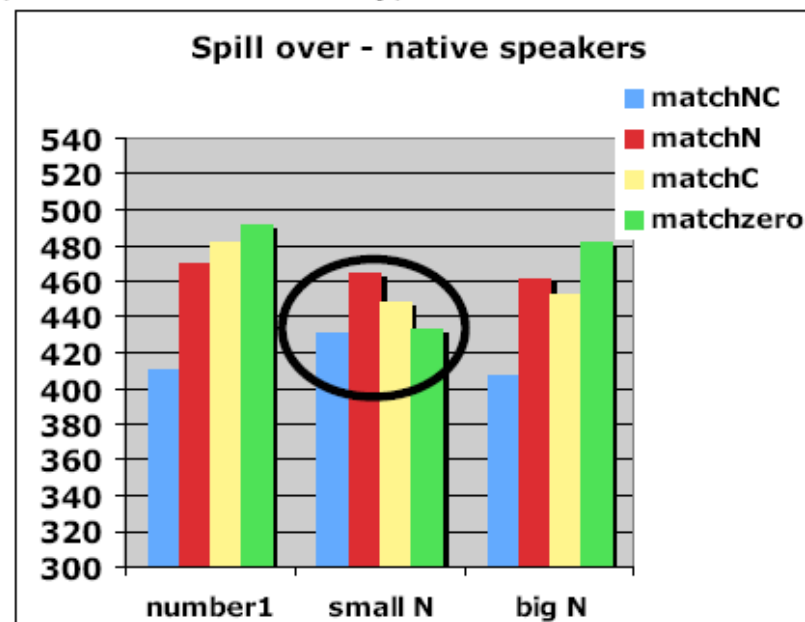
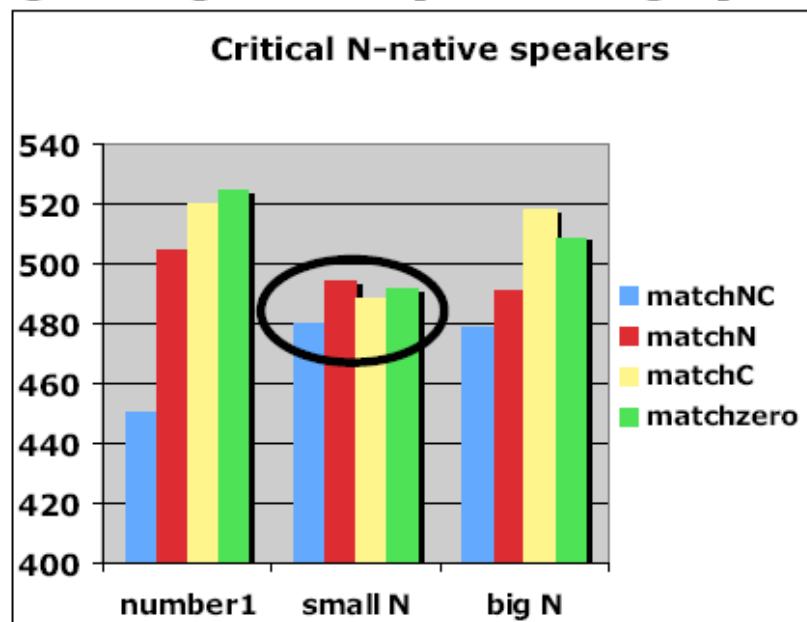
- Subjects containing paucal numerals can trigger plural agreement (rather than masculine/feminine singular) and control plural anaphora:

(10) Tri mal'čhika so smexom gonjajutsja drug za drugom.
three boy."GEN.SG" with laughter run.3.PL one-after-another.RECIP
'Three boys run after one another with laughter.' (NCR)

“GEN.SG” is not processed as genitive singular!

- Xiang, Harizanov & Polinsky (2008a, b): “GEN.SG.” does not receive the representation as a singular genitive – otherwise subjects would have shown sensitivity to the wrong representations

Fig. 1: Xiang et al's Self-paced reading experiment (masculine nouns only):



For paucal numerals:

- match NC = GEN.SG
- match N = NOM.SG
- match C = GEN.PL
- match zero = NOM.PL

- Paucal numerals present the biggest problem for human parsers
- As far as Spill over time is concerned, the most acceptable forms of nouns are GEN.SG and **NOM.PL** (match zero!)

Paucal numerals usage in Russian

- Total 105 speakers
 - not all answered all questions
 - 39 male, 66 female
 - ages 18-75
 - 76 non-linguists
- Two-part questionnaire:
 - grammaticality judgments: 'Is this sentence acceptable for you?', 'Can this phrase refer to a group of males? females? mixed group?'
 - preference test: fill in the blank
blah-blah _____(x, y, z) blah-blah

Feminine de-adjectival nouns

(11) В доме лорда Паркера было много прислуги: дворецкий, экономка, конюхи и три _____ (горничной, горничных, горничные).

'In Lord Parker's house there were many servants: a butler, a housekeeper, some horsegrooms and three _____ (maids: GEN.SG, GEN.PL, NOM.PL)

| form | speakers prefer | percentage | |
|-----------------------|-----------------|------------|----|
| gorničnye (NOM.PL) | 48/85 | 56.5% | ok |
| gorničnyx (GEN.PL) | 37/85 | 43.5% | % |
| gorničnoj (GEN.SG) | 0 | 0% | * |

- Cf. also Mel'čuk (1985: 176-177)
- National Corpus of Russian: a study of 24 such nouns shows an average 5:1 ratio of occurrence for NOM.PL vs. GEN.PL (for *gorničnaja* 'maid' the ratio of occurrences is 10:1)

Feminine last names in *-ova*

(12) a. В нашем классе -- три Ивановы: Женя, Валя и Саша.

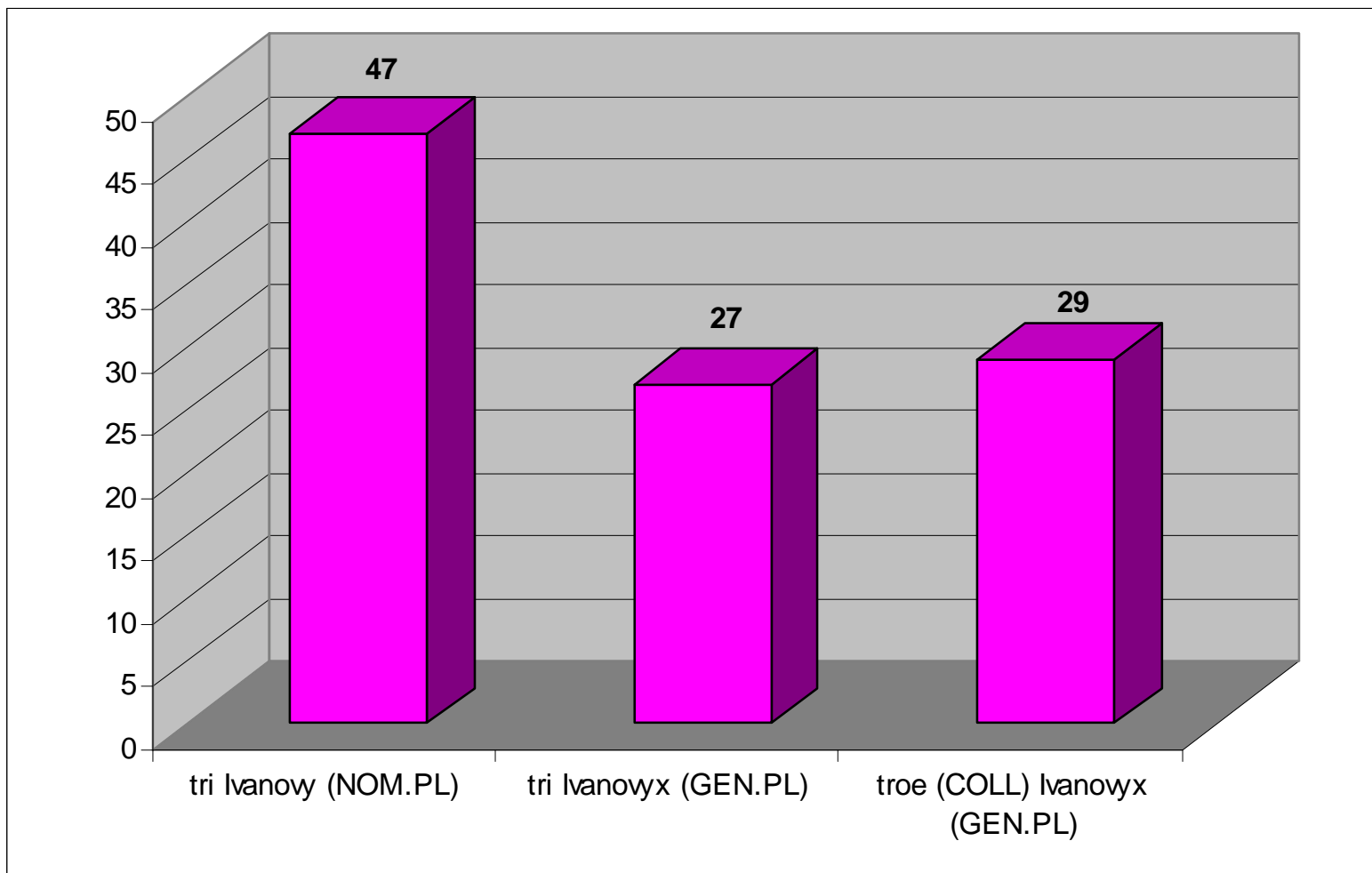
'In our house there are three Ivanovs (**NOM.PL**): Zhenya, Valya and Sasha.'

b. В нашем классе -- три Ивановых: Женя, Валя и Саша.

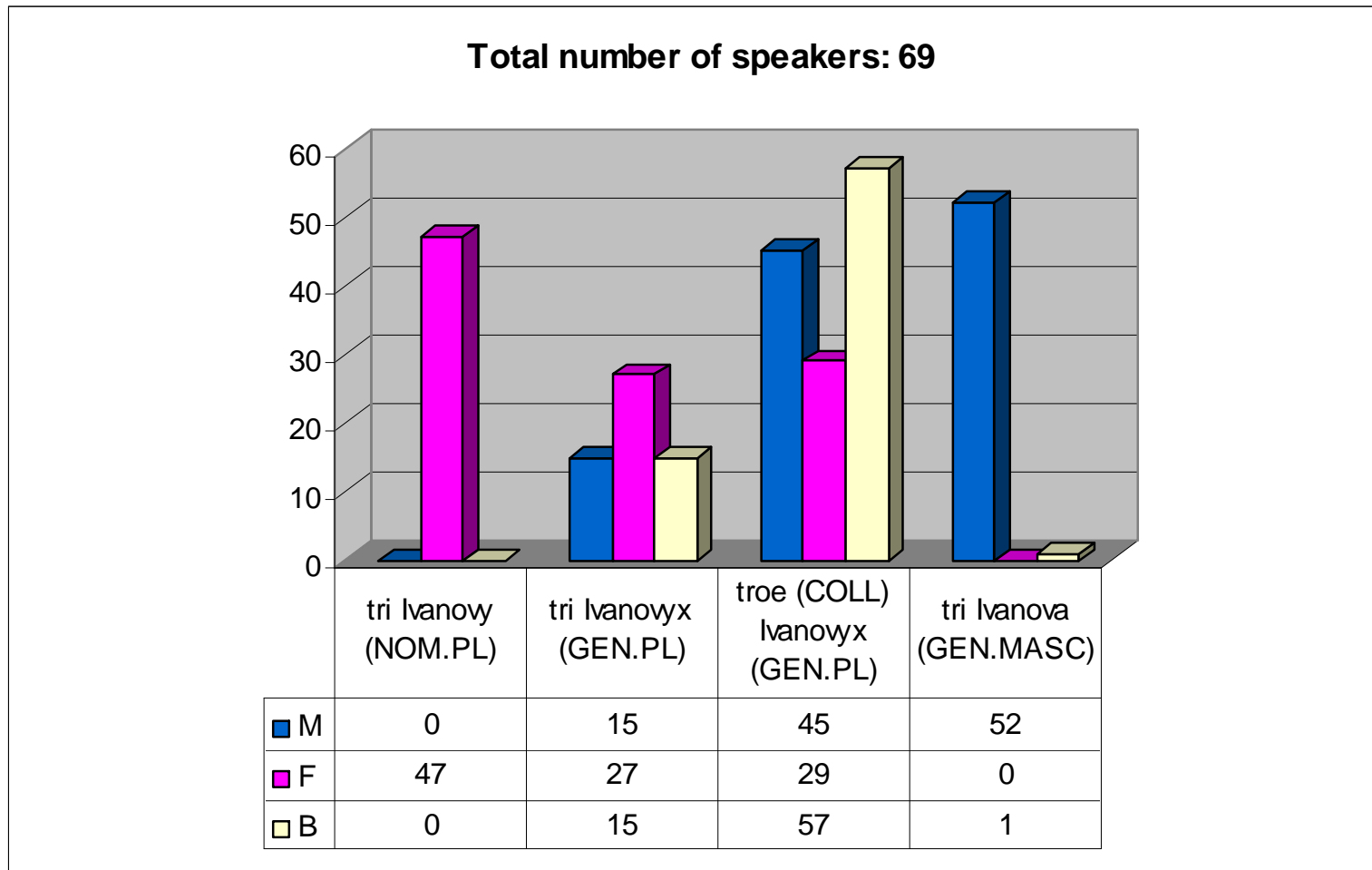
'In our house there are three Ivanovs (**GEN.PL**): Zhenya, Valya and Sasha.'

| form | speakers accept | percentage | |
|----------------------|-----------------|------------|----|
| Ivanovy (NOM.PL) | 47/69 | 68.1% | ok |
| Ivanovyx (GEN.PL) | 27/69 | 39.1% | % |
| Ivanova (GEN.SG) | 0 | 0% | * |

Speakers' acceptance of case/number forms to refer to a group of females



Speakers' acceptance of case/number forms to refer to a group of males, a group of females or a mixed group



Adjectives modifying feminine nouns

(13) Список его произведений включает в себя четыре _____ (детских, детские, детской) книги, десяток коротких рассказов и несколько пьес.

'The list of his works includes four _____ (children's (Adj): GEN.PL, NOM.PL, GEN.SG) books, ten short stories and several plays.'

| form | speakers prefer | percentage | |
|---------------------|-----------------|------------|----|
| detskie (NOM.PL) | 61/86 | 70.9% | ok |
| detskix (GEN.PL) | 25/86 | 29.1% | % |
| detskoj (GEN.SG) | 0 | 0% | * |

- Cf. also Graudina et al. (1976: 32-33), Mel'čuk (1985: 126-128); Gramota.ru (#240687, 211404, 216078)
- Once again, corpora study indicates a strong preference for NOM.PL over GEN.PL

Adjectives modifying stress-distinguishing feminine nouns

- *gorý* = GEN.SG, *góry* = NOM.PL
- Grammatical descriptions (cf. Graudina et al. 1976: 32-33, Mel'čuk 1985: 126-128) state that with such stress-distinguishing nouns adjectives appear in GEN.PL
- However, my corpora searches show that NOM.PL is preferred for all types of nouns

(14) Он окинул взглядом открывающийся перед ним живописный пейзаж: две _____ (высокие, высокой, высоких) горы и журчащая между ними речка.

'He looked at the beautiful landscape in front of him: two _____ (tall: NOM.PL, GEN.SG, GEN.PL) mountains and a river gurgling between them.'

| form | speakers prefer | percentage | |
|---------------------|-----------------|------------|----|
| высокие (NOM.PL) | 62/86 | 72.1% | ok |
| высоких (GEN.PL) | 24/86 | 27.9% | % |
| высокой (GEN.SG) | 0 | 0% | * |

Adjectives modifying stress-distinguishing feminine nouns (2)

- Only 12/85 speakers' (14.1%) responses are consistent with Graudina et al.'s claims
- ... while 13/85 speakers' (15.3%) responses are the opposite: they prefer NOM.PL on adjectives modifying stress-distinguishing nouns and GEN.PL on other adjectives

Are the responses consistent within a category?

- Consistent treatment of adjectives (regardless of what nouns they modify): 60/85 speakers (70.6%)
- Consistent treatment of nouns (*gorničnaja*-type and last names in *-ova*): 28/49 speakers (57.1%)
- Treat *gorničnaja*-type nouns as Adjectives: 41/84 speakers (48.8%)



Interim conclusion

- The apparent “GEN.SG” is neither genitive, nor singular
- If so, what is it?



Proposal

- The apparent “GEN.SG” on complements of paucal numerals is **Quantificational Paucal**

How it can be done?

- Distinguish abstract syntactic Cases and their morphological realizations: a given m-case may spell-out a number of distinct syntactic Cases, and conversely a given syntactic Case may be spelled-out by a number of distinct m-cases
- The Case checked in the heterogeneous case distribution is **QUANT** (Quantificational) Case
 - It can be realized as GEN, NOM, or a special “count form”.
 - cf. Mel’čuk (1985: 435), Isakadze (1998: 85-86), Rakhlin (2003), Rappaport (2004)
 - for Bailyn (2004): QUANT = Case of Q, or uQ on N

(15) a. pjat’ knig b. tri knigi
 [QUANT] [QUANT] [QUANT] [QUANT]

Why different forms for [QUANT]?

- Lexical Insertion operates on bundles of features, incl. Case and Num
 - And also categorial features, gender, animacy (and for nouns, declension class)
 - Lexical Insertion rules are subject to dialectal and diachronic variation
- The difference between *knig* and *knigi* in (15) has to do with the number feature: [PL] vs. [PAUC]:
 - cf. Isakadze (1998: 85-86), Pesetsky (2007), Rakhlin (2003), Bailyn & Nevins (2008)
 - for Rappaport (2004), PAUC is a Case, not a number feature

| | | | | | |
|---------|---------|---------|----|---------|---------|
| (16) a. | pjat' | knig | b. | tri | knigi |
| | [QUANT] | [QUANT] | | [QUANT] | [QUANT] |
| | [PL] | [PL] | | [PAUC] | [PAUC] |

What is paucal number?

(17) Bailyn & Nevins (2008):

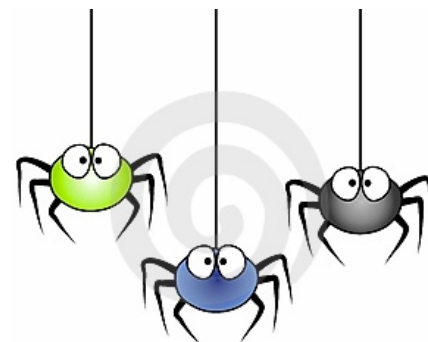
Singular: [+singular]

Paucal: [−singular, −augmented]

Plural: [−singular]

- Verbal agreement expresses only [±singular]; hence, paucal noun phrases trigger plural agreement; cf. (10)

«паукальное число»? →



Plurals and gender

- Plurals (=non-singulars) are specified for gender; it is not expressed on plural forms because of impoverishment/neutralization:
 - cf. Zaliznjak (1964), Bailyn & Nevins (2008):

(18) tri vysokie devuški,
 každaja iz kotoryx...

three tall.NOM.PL girls.“GEN.SG”
each.**FEM** from which.GEN.PL

‘three girls each of which...’

The old question: which is the head: the numeral or the noun?

- My proposal: BOTH!
- Some features are specified on the noun (and the numeral agrees), while other features are specified on the numeral (and the noun agrees):
 - Noun/N^o is specified for [Gender]
 - Numeral/Q^o is specified for [Case: QUANT] and [Number]

Consequences: number-less plurals

- Unless Q^o is merged, nouns are number-less. The default spell-out of number in Russian is PL:
 - Cf. Pereltsvaig (2006a), Pesetsky (2007), Kagan & Pereltsvaig (2009)
- (19) Terminator sobralsja v kalifornijskie gubernatory.
Terminator prepared into Californian.PL governor.**PL**
'The Terminator prepared (to run) for the California governor.' (Google)
- (20) remont {komp'juterov / avtomobilej }
repair computer.**PL** / cars.**PL**
'computer repair/ auto repair'
- (21) Lena najelas' [NP kotlet].
Lena *na-ate-sja* burger.GEN.**PL**
'Lena ate her fill of burgers.'

Consequences: Pluralia tantum

- *Pluralia tantum* nouns are specified for number [–singular]; if Q° is specified as [–singular, –augmented], neither agrees with the other and we get a clash:

(22) a. * {dva / tri / četyre} {brjuk / očkov / nožnic}
{two.PAUC / three.PAUC / four.PAUC} {pants / glasses / scissors}

b. {dvoe / troe / četvero} {brjuk / očkov / nožnic}
{two.COLL / three. COLL / four. COLL} {pants / glasses / scissors}
'{two/three/four} pants, glasses, scissors'

- The requirement for collective numerals with *pluralia tantum* nouns applies only to paucal numerals:

(23) {pjat' / sem' / desjat'} {brjuk / očkov / nožnic}
{five.PL / seven.PL / ten.PL} {pants / glasses / scissors}
'{five/seven/ten} pants, glasses, scissors'

Example:

(*italics* for valuation through agreement; **boldface** for feature values expressed morphologically)

| | | | |
|----------|------------|--------------|--------------|
| | dve | interestnye | knigi |
| category | Q | A | N |
| Case: | QUANT | <i>QUANT</i> | <i>QUANT</i> |
| Gender: | <i>FEM</i> | <i>FEM</i> | FEM |
| Number: | PAUC | <i>PAUC</i> | <i>PAUC</i> |
| ... | | | |

Adjectives modifying numerals (Babby's "prequantifiers")

- Typically, these occur in the "GEN.PL":

(24) celyx pjat' dokladov pro čislitel'nye
whole."GEN.PL" five talks about numerals
'a whole five talks on numerals'

- The same is true with paucal numerals:

(25) celyx tri doklada pro čislitel'nye
whole."GEN.PL" three talks about numerals
'a whole three talks on numerals'

Problem:

- If this "GEN.PL" is really QUANT.PAUC, why is it spelled-out as -yx rather than -ye with feminine nouns?
- Can [A, QUANT, PAUC, FEM] be spelled-out in two different ways depending on the adjective?

(26) celyx tri novye knigi
 whole."GEN.PL" three new."NOM.PL" books
 `a whole three new books`

- 61/86 speakers (70.9%) prefer this over *celye... novye*

Interesting facts:

- spread of “NOM.PL” -ye as spell-out for [A, QUANT, PAUC, FEM]:

(27) Он опубликовал _____ (целой, целые, целых) три книги по синтаксису русских числительных.

‘He published _____ (whole: GEN.SG, NOM.PL, GEN.PL) three books on the syntax of Russian numerals.’

(28) У меня есть для вас _____ (целой, целые, целых) две хорошие новости и только одна плохая.

‘I have for you _____ (whole: GEN.SG, NOM.PL, GEN.PL) two good (NOM.PL) pieces of news and only one bad one.’

| form | speakers prefer | percentage |
|---|-----------------|------------|
| celye (NOM.PL) 3 knigi | 5/85 | 5.9% |
| celye (NOM.PL) 2 хороšie (NOM.PL) novosti | 25/85 | 29.4% |

What is going on with prequantifiers?

- Possible cause for the change: discriminating adjectives modifying FEM and MASC nouns:
 - contra Borrás & Christian (1971), Graudina et al. (1976), Bailyn & Nevis (2008)

(29) dva spasitel'nye majaka ← archaic
two saving."NOM.PL" lighthouse."GEN.SG"
'two saving lighthouses' (M. Lermontov, *Geroj našego vremeni* (1838-1840))

- **MASC**: "NOM.PL" -ye → "GEN.PL" -yx; cf. (29)
- **FEM**: "GEN.PL" -yx → "NOM.PL" -ye (younger speakers prefer the latter)
- **prequantifiers**: there is a shift too, esp. in the presence of a "NOM.PL" adjective modifying N



CONCLUSION

- Babby's contrast between heterogeneous and homogeneous patterns is only partially syntactic (merger of numeral in head or Spec; QUANT Case), while the complications in the realization of the QUANT Case are to be accounted for in the morphology.



Thank you:

- John Bailyn
- Gabi Danon
- Olga Kagan
- Saya Kitasei
- Jim Lavine
- Maria Polinsky
- Gil Rappaport
- Ming Xiang
- All the consultants from my empirical study pool, too numerous to be listed here individually