

Optional number agreement in Bashkir (a micro-corpus study)

1. The phenomenon and aims of the study

- Predicate number agreement with 3rd person subjects is optional in Bashkir: (semantic) plurality of subjects is a necessary, but not a sufficient condition for plural marking of the predicate:

- (1) *bala-lar* *uqə-j* / *uqə-j-ǰar*
child-PL learn-PRS / learn-PRS-PL
'Children learn.' [Юлдашев (ред.) 1981: 252]

- The goal of the study: to unveil those factors that determine the (probability of the) presence / lack of the plural marking of the predicate.

2. Data

- The data were obtained from a micro-corpus that contains semi-spontaneous oral texts. The texts were recorded and analyzed collectively (many thanks to everybody!) during fieldwork trips to Bashkortostan (2011–2014). Its total size is about 1200 “sentences” (≈140 min.).

- We extracted **all** clauses that conformed to the following conditions:

- finite clause (predicate agreement is impossible or different in non-finite clauses), irregardless of whether the predicate is verbal or non-verbal;
- 3rd person subject;
- semantically plural subjects, including collective (“corporate” [Corbett 2000: 188]) nouns:

- (2) *milicija* *kil-de,* *tikšer-ǰe-lär*
police come-PST verify-PST-PL
'Police came and made an investigation' (two clauses).

- Altogether **305** clauses were extracted: **230** (75%) had a plural marker, **75** (25%) lacked it.

3. Data annotation

- Each clause was viewed as a potential locus of agreement, e.g. (2) has two loci.
- Each clause was annotated for 10 parameters:
 - Presence/absence of the plural morpheme on the predicate (dependent variable)
 - Impersonal interpretation: YES / NO
 - Word order: SV, SXV, VS, VXS
 - Animacy of the subject: human / animate / inanimate
 - Referentiality of the subject: definite / specific (referential) indefinite / non-referential
 - Givenness of the subject: given (mentioned previously in discourse) / new

- Syntactic structure of the subject: plural NP, conjoined NPs, comitative NP, zero, etc.
 - Presence of the following same-subject finite clause within the same sentence: YES / NO
 - Semantics of the predicate: joint / distributive action (cf. [Kemmer 1993])
 - ⊗ contrary to some previous claims [Дмитриев 2008: 206; Poppe 1964: 91–93] this factor appeared to be irrelevant in our data; it is not discussed below.
 - Syntactic structure of the predicate: various verb forms, non-verbal predicates.
- Why these parameters?
- occasional claims with respect to their relevance for agreement in Bashkir, typically without much empirical evidence [Дмитриев 2008; Юлдашев 1981; Poppe 1964];
 - studies on optional number agreement in other languages [Corbett 2000, 2006; Creissels 2010; Durie 1986; Iemmolo (in press); Пекелис 2013];
 - our hypotheses that emerged when collecting the data.

4. Results

- In sections 4.1–4.3 we discuss stronger factors, in some cases their impact borders exceptionlessness. Hence, some strongly predictable clauses are excluded from the analysis in sections 4.4–4.7.

4.1. Impersonal constructions

- Indefinite / generic type, «R-impersonals» in terms of [Malchukov & Ogawa 2011: 44ff.]
- ⇒ Expectation: [Юлдашев (ред.) 1981: 437] claims that in such clauses («неопределенно-личные») the predicate is always in the 3PL form.
- ⊗ Result: the vast majority of predicates in impersonal constructions do have a plural marker (see ex. (3), and figures in Table 1), but the pattern is not exceptionless, see (4).

(3) *serial-ə-n ike tapqər kür-hät-te-lär*
 serial-P.3-ACC two times see-CAUS-PST-PL
 ‘(They) showed this serial 3 times’ (≈ ‘this serial was shown on TV 3 times’).

(4) *Äwžän-gä jebär-ä tor-yan ine beð-ðe elek*
 Avzjan-DAT send-CV.IPFV stand-PC.PST be.PST we-ACC formerly
 ‘Formerly, (they) used to send us to Avzjan’ (≈ ‘we used to be sent to Avzjan’).

Table 1. Impersonal clauses and plural marking of the predicate

	–PL		+PL		Total
	N	%	N	%	N
Impersonal clauses	6	6%	96	94%	102
Other clauses	69	34%	134	66%	203
TOTAL	75	25%	230	75%	305

- In such sentences plural marker may not play the role of the agreement marker, but rather may function as a marker of the impersonal indefinite-subject clause type as such.
- Below, impersonal clauses will not be taken into account. Only **203** non-impersonal clauses will be analyzed the sections to follow.

4.2. Syntactic structure of the predicate

Table 2. Syntactic structure of the predicate and plural marking of the predicate

Predicate type:	-PL		+PL		Total
	N	%	N	%	N
<i>bar</i> ‘there is’ / <i>juq</i> ‘there isn’t’	4		0		4
PC.PST	16	35%	30	65%	46
PST	9	30%	21	70%	30
PRS	29	28%	75	72%	104
Other	11		8		19
TOTAL	69	34%	134	66%	203

- No significant differences between most common types of predicates (PC.PST, PRS, PST).
- Although the data are scarce, existential predicates *bar* ‘there is’ / *juq* ‘there is not’ strongly disfavour plural marking of the predicate (possibly block it). These are also the only clauses with non-referential subjects in our sample (apart from impersonal clauses, see above).

(5) *ike balam bar*
 two child-P.1SG there.is
 ‘I have two children.’

- Thus, only remaining **199** clauses are further analyzed.

4.3. Animacy of the subject

- ⇒ There is a correlation between animacy of the subject and agreement [Corbett 2006: 177–179, 191; Пекелис 2013]; also mentioned in passing for Bashkir [Дмитриев 2008: 206].
- In clauses with non-human subjects (14 inanimate, 3 animate non-human) the predicate never has a plural marker (in our corpus):

(6) *tör-lö versija-lar başla-n-a*
 variety-ADJ version-PL begin-REFL-PRS
 ‘Different versions emerge’.

Table 3. Animacy hierarchy and plural marking of the predicate

	-PL		+PL		Total
	N	%	N	%	N
inanimate	14		0		
animate non-human	3		0		
human	48	26%	134	74%	182
Total	65	33%	134	67%	199

- Thus, there is a strong correlation between non-humanness of the subject and absence of agreement. Only 182 clauses with animate subjects are analyzed in the remainder of the talk.

4.4. Givenness and referentiality of the subject

- ⇒ Given subjects are known to be more likely to trigger agreement. Agreement suspension is characteristic of existential or presentational constructions [Corbett 2006: 197 ff; Creissels 2010; Iemmolo (in press)].

- ⇒ The higher the referential status of the subject, the likelier is number agreement [Corbett 2006, 200 ff; Konoshenko 2010].
- ⊗ Our data strongly support both predictions: “given” subjects trigger agreement more often than “new” subjects, definite subjects trigger agreement more often than indefinite subjects (χ^2 , $p < 0.01$ in both cases).

Table 4. Given/new subject and plural marking of the predicate

	-PL		+PL		Total
	N	%	N	%	N
given	22	17%	104	83%	126
new	26	46%	30	54%	56
TOTAL	48	26%	134	74%	182

Table 5. Definiteness of the subject and plural marking of the predicate

	-PL		+PL		Total
	N	%	N	%	N
definite	23	20%	94	80%	117
indefinite	25	38%	40	62%	65
TOTAL	48	26%	134	74%	182

- These two parameters are closely interconnected. Probably (cf. Tables 4 and 5) givenness is a stronger factor than definiteness.

4.5. Syntactic type of subject: implicit vs. explicit

- ⇒ In Turkish, number agreement is compulsory if the subject is implicit [Corbett 2006, 190].
- ⊗ In our data, predicates in clauses with implicit (semantically plural) subjects are almost invariably marked for number (cf. Table 6):

(7) *ađaq zəjan-də qapla-nə-lar inde*
 then harm-ACC cover-PST-PL yet

{Context: criminals were captured and they have admitted their guilt.} ‘Then, they have repaired the damage’

- Implicit (“zero”) subjects can reflect the highest degree of the referent’s activation in discourse [Givón 1995: 379]. Hence, when comparing implicit and explicit subjects we present data only for given subjects:

Table 6. The expression of subject and plural marking of the predicate (only for “given” subjects)

	-PL		+PL		Total
	N	%	N	%	N
“Zero” (implicit subject)	2	4%	49	96%	51
explicit subject	20	27%	55	74%	75
TOTAL	22	17%	104	83%	126

- ⊗ Even for “given” subjects, implicitness (non-overtness) of the subject favors plural marking of the predicate (χ^2 , $p < 0.001$).

4.6. Presence of the following same-subject finite clause within the same sentence

⇒ We noticed that finite sentences with chaining are unusual with respect to agreement:

- (8) *jäš-tār uram-da jörö-j, es-ä-lär, huy-əš-a-lar*
 young-PL street-LOC go-PRS drink-PRS-PL hit-RECP-PRS-PL
 ‘Youngsters walk, drink, fight.’

- There were 11 sentences in which the first predicate did not have a plural marker, while the second one did. There were no sentences with the opposite pattern.
- ✎ Initial predicate in a chaining structure is the position that disfavors plural marking.

4.7. Word order

- ⇒ According to typological literature, agreement is more likely for subjects preceding predicates [Corbett 2006: 180; Creissels 2010; Iemmolo (in press); Пекелис 2013].
- ⇒ Corbett [2000: 210] mentions that in Cairene Arabic “real distance” had an important role: in structures of the VXS type agreement is more likely than in structures of the VS type:

Table 7. Effect of distance of target from controller in Cairene Arabic [Corbett 2000: 210]

Distance from head (in words)	Plural (semantic) agreement %	N
1	21	276
2	36	115
3-5	43	141
6-8	47	57
9-45	91	64

Table 8. Word order and plural marking of the predicate (Bashkir data)¹

	-PL		+PL		Total
	N	%	N	%	N
SXV	17	20%	69	80%	86
SV	23	82%	5	18%	28
VS	2		2		4
VXS	1		2		3
TOTAL	43	36%	78	64%	121

- ✎ Insufficient data for making any conclusion with respect to preverbal (SV, SXV) vs. postverbal (VS, VXS) subjects (but NB: postverbal subjects can trigger agreement).
- ✎ There is a striking difference between SV (disfavor agreement, cf. (9)) and SXV (favor agreement, cf. (10)) structures. The difference is statistically significant (χ^2 , $p < 0.001$).

- (9) *hat-əp al-əw-sə-lar kil-de*
 sell-CV take-NMLZ-AG-PL come-PST
 ‘Customers came.’

- (10) *unda jäšä-gän bašqort-tar =əa küb-eräk rus-sa höjlä-š-ä-lär*
 there live-PC.PST bashkir-PL PCL much-CMPR russian-ADV speak-RECP-PRS-PL
 ‘Bashkir people who live there mostly speak Russian.’

¹ Clauses with implicit subjects and 7 problematic cases were disregarded, hence, only 121 clauses overall.

- In order to partial out the potential impact of the givenness parameter, the data were tabulated separately for given (Table 9) and new subjects. The difference is statistically significant (Fisher’s exact test, $p < 0.001$ in the former case, ≈ 0.05 in the latter case).

Table 9. Word order and plural marking of the predicate (“given” subjects only)

	-PL		+PL		Total
	N	%	N	%	N
SXV	10		50		60
SV	7		2		9
TOTAL	17		52		69

- The likelihood of plural marking on the predicate is drastically higher if the subject and the predicate are non-adjacent.

5. Summary and conclusions

Table 10. Factors that influence the presence /lack of the plural marking of the predicate (rule-like generalizations are boldfaced)

Parameters	Values that correlate with...		
	+PL	-PL	generalization
Subject: animacy		non-human	PL is favoured if subject is salient
Subjects: “givenness”	given	new	
Subjects: definiteness	definite	indefinite	
Predicate: structure		existential (<i>bar/juq</i>)	
Impersonal constructions	impersonal		PL is disfavoured if its use creates redundancy
(Given) subjects: zero/ explicit	zero	explicit	
Distance between subject and predicate	non-adjacent (SVX)	adjacent (SV)	
Chaining structures		initial position	

- The likelihood of the plural marking of the predicate is higher for **more salient subjects** (animate, definite, given), cf. the gray zone in Table 10. This regularity perfectly corresponds to typological expectations.
- Plural marking of the predicate occurs more frequently if **it does not lead to grammatical redundancy**, i.e. in clauses without an overt subject (either indefinite / generic or definite / topical). Constructions in which the idea of plurality is expressed more than once are more likely if the sites of coding plurality (subject and predicate) are at a greater distance from each other. This tendency can be also related to the tendency to avoid redundancy (“density of redundancy”).
- Therefore, the verbal plural marker in Bashkir can be viewed not as a unit that copies grammatical features of the subject, but rather as a marker which is relatively independent from the nominal plurality, and which partially tends to appear in complimentary distribution with markers of nominal plurality.

References

- Дмитриев Н.К. 2008. Грамматика башкирского языка. (Переизд.). Москва: Наука.
- Падучева Е.В. 2012. Неопределенно-личное предложение и его подразумеваемый субъект // Вопросы языкознания, № 1. С. 27–41.
- Пекелис О. Е. 2013. "Частичное согласование" в конструкции с повторяющимся союзом: корпусное исследование основных закономерностей // Вопросы языкознания, № 4. С. 55–86.
- Плунгян В.А. 2009. Общая морфология. Издание 3-е. М.: УРСС.
- Юлдашев А. А. 1981. (ред.). Грамматика современного башкирского литературного языка. М.: Наука.
- Corbett G. G. 1992. A typology of number systems. European Science Foundation Programme in Language Typology: Theme 7, Noun Phrase Structure: Working Paper no. 15.
- Corbett G. G. 2000. Number. Cambridge: CUP.
- Corbett G.G. 2006. Agreement. Cambridge: CUP.
- Creissels D. 2010. Fluid intransitivity in Romance languages // Archivio glottologico italiano, Vol. 95, № 2. P. 117–121.
- Durie M. 1986. The grammaticization of number as a verbal category // V. Nikiforidou et al. (eds.). Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society. Berkeley: University of California. P. 355–370.
- Givón T. 1995. Functionalism and grammar. Amsterdam/Philadelphia: John Benjamins.
- Iemmolo G. (in press). Subject agreement suspension from Old to Modern Tuscan: between syntax, morphology and information structure.
- Kemmer S. 1993. Marking oppositions in verbal and nominal collectives // *Faits de langues*, № 2. P. 85–95.
- Konoshenko M. 2010. Person-Number Agreement in Mande: Synchronic Hierarchies and Their Sources. Paper presented at the biennial conference of the Association for Linguistic Typology (ALT10). Leipzig.
- Levinson S.C. Pragmatics. Cambridge: CUP. 1983.
- Malchukov, Andrej & Akio Ogawa. 2011. Towards a typology of impersonal constructions: A semantic map approach // A. Malchukov, A. Siewierska (eds.). Impersonal constructions: A cross-linguistic perspective. Amsterdam / Philadelphia: John Benjamins.
- Poppe N. 1964. Bashkir manual. Bloomington: Indiana University.