

# Preposition-drop in contact-influenced Russian Speech of the Russian Far East and Northern Siberia

Irina Khomchenkova<sup>1,2,3</sup>, Natalya Stoyanova<sup>1,2</sup>

1 Vinogradov Russian Language Institute, RAS;

2 Institute of Linguistics, RAS;

3 Lomonosov Moscow State University

# Introduction

- The paper deals with a feature of Russian varieties used by speakers of indigenous languages (IL) of Russia, referred to as “**preposition-drop**” (1).
- The preposition (in angular brackets) is absent, and the noun demonstrates the same case form as in the corresponding Monolingual Russian PP.

(1) *Ja ži-l-a*                      <**v**> *Jakolevk-e*.  
I live-PST-SG.F            **in** Yakolevka-**PREP**  
'I lived in Yakolevka.'

# Data: texts

- The data of two bilingual Russian varieties featuring the variation between standard PPs and structures with preposition-drop.
  - **“Tungusic Russian”**: a variety used in the Amur region by speakers of Southern Tungusic languages (Nanai, Ulcha);
  - **“Samoyedic Russian”**: a variety used in the Taimyr peninsula by speakers of Northern Samoyedic languages (Tundra Enets, Forest Enets, Nganasan).
- **Our data: the corpus of Contact-influenced Russian Speech** (<http://web-corpora.net/wsgi3/ruscontact/search>), created by our team:
  - ca. 48,000 tokens for Tungusic Russian;
  - ca. 47,000 tokens for Samoyedic Russian.

# Data: PPs

- The total sample of “PPs”:
  - only those prepositions that are attested to be omitted: *v* ‘in’, *k* ‘to’, *s* ‘from, with’, *na* ‘on(to)’, *u* ‘near’.

	Tungusic Russian			Samoyedic Russian		
		omitted	overt		omitted	overt
<i>v(o)</i> ‘in’	49,54%	268	273	22,76%	127	431
<i>s(o)</i> ‘from, with’	16,55%	46	232	3,38%	9	257
<i>k(o)</i> ‘to’	14,94%	13	74	13,27%	13	85
<i>na</i> ‘on’	10,11%	55	489	0%	0	541
<i>u</i> ‘at’	1,69%	8	465	0%	0	537
total	<b>20,28%</b>	<b>390</b>	<b>1533</b>	<b>7,33%</b>	<b>147</b>	<b>1858</b>

# Hypotheses on preposition-drop

- **Phonetic influence hypothesis:**
  - P-drop emerges as a cluster-avoiding device under the influence of the IL, which has restrictions on word-initial clusters.
  - Tungusic & Samoyedic have no initial clusters and no exact equivalents to the Russian [v], [f], which are the phonetic realizations of the preposition v 'in'.
- **Morphosyntactic influence hypothesis:**
  - P-drop is a result of the copying of morphosyntactic patterns of the IL.
  - Tungusic & Samoyedic have spatial cases, postpositions, but no prepositions.
- **Non-contact hypothesis:**
  - P-drop is regulated by principles of a non-contact nature, such as those determining the choice between more explicit vs. reduced means of expression cross-linguistically.

# Analysis

- A multifactorial statistical analysis of the most frequently omitted preposition *v* 'in'
  - the logistic regression model with mixed-effects, speaker as a random effect
- A preliminary analysis:
  - sociolinguistic parameters, inter-speaker variation
  - omission of prepositions other than *v* 'in'

# The omission of *v* 'in': predictors

- **phonetic context** [consonant / vowel] **Phonetic H**
  - *v* may drop more easily before consonants, since consonant clusters break the phonological rules of Tungusic and Samoyedic languages.
- **case** [PREP = Location (*v tundr-e* 'in tundra') / ACC = Goal (*v tundr-u* 'to tundra')] **Morphosyntactic H**
  - *v* may drop more easily before PREP, since
    - In Tungusic PREP corresponds to case forms, and PPs with ACC correspond to PostP (for TungRus)
    - PREP is used only with prepositions → the preposition is redundant
- **noun semantics** [time / place / other] **Non-contact H**
  - *v* may drop more easily in PPs referring to place (toponyms, topographical objects...) and time names, in which the preposition is less informative & which tend to be expressed in a more reduced way cross-linguistically

# The omission of *v* 'in': results

	<b>Tungusic Russian</b>	<b>Samoyedic Russian</b>	<b>Hypothesis</b>
<b>Phonetic context</b>	before consonants	no effect	Phonetic: only for Tungusic Russian
<b>Case</b>	ACC (Goal)	no effect	Morphosyntactic & Non-contact: ???
<b>Noun semantics</b>	place & time names		Non-contact: yes

# The omission of other prepositions

- **Non-syllabic vs. syllabic** forms of “short” prepositions:  
*v* ‘in’, *k* ‘to’, *s* ‘from, with’ ~ *vo*, *ko*, *so*  
(*s polki* ‘from the shelf’ ~ *so stola* ‘from the table’)

Tungusic Russian: the omission of non-syllabic variants is significantly more often

→ The phonetic factor is at play.

- **“Long” prepositions** (open-syllable): *na* ‘on’, *u* ‘near’ - are attested to be omitted in Tungusic Russian, although with a low frequency
- **S ‘with’ vs. s ‘from’**: in Tungusic Russian, the omission of *s* ‘with’ is significantly more often  
→ The phonetic factor is not the only one.

# Sociolinguistic factors & inter-speaker variation

- **Year of birth:** no clear correlation
- **Level of education:** no clear correlation
- **L1 || area:** preposition-drop is more frequent in the Tungusic sample than in the Samoyedic one
- Speakers who avoid word-internal clusters: no clear correlation  
→ Contra: phonetic hypothesis
- Speakers who have more contact-induced syntactic features in their Russian speech: drop prepositions more often (only for Tungusic Russian)  
→ Pro: morphosyntactic hypothesis

# Discussion: different factors in interplay

universal semantic factors



Receiving  
Russian input

more information on  
more frequent contexts  
(e.g. with place-names)

*v shkole* 'at school'

phonetic influence



Analyzing  
Russian input

basing on L1 phonetic  
keys: frequent PPs with  
clusters are  
reinterpreted as bare  
case forms

*v shkole* 'at school'

morphosyntactic  
influence (partly)



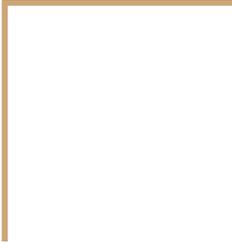
Producing  
(non-standard)  
Russian speech

bare case forms are  
overgeneralized to other  
contexts: especially to  
those supported by  
polysemy models of L1

~~na~~ *rabote* 'at work' (LOC=DAT)

# Conclusion

- Preposition-drop is caused by the interplay of:
  - general non-contact semantic factors;
  - the phonetic influence of indigenous languages;
  - morphosyntactic calquing (?).
- The results are much more clear for the Tungusic sample, in which preposition drop is more widely attested than in the Samoyedic one.



# Details





# 1. Introduction: preposition drop



# Preposition drop and similar features

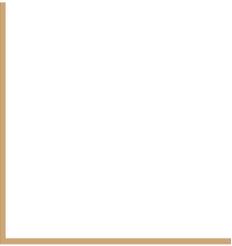
- **Preposition drop** (the use of noun with vs. without preposition in the same context) is attested in different languages. It occurs both in contact-influenced varieties and across monolinguals.
- For instance:
  - different varieties of English (Myler 2013; Bailey 2018), Italian dialects (Cattaneo 2009: 286–290), Modern Greek (Ioannidou & Den Dikken 2009; Terzi 2010; Gehrke & Lekakou 2012; Sifaki 2020).
- Different syntactic explanations were proposed for structures without prepositions in different languages.
- Different contexts were reported to be prone to preposition drop, e.g.:
  - Goal-contexts with motion verbs
  - Place-names: toponyms, institutions, ...
- A more general discussion on **zero-marking and other asymmetries in encoding of spatial arguments**, cf. (Stolz et al. 2014; Haspelmath 2019):
  - “place names” are more often encoded in a more reduced way
  - “deviations from usual associations of role meanings and properties of referring expressions tend to be coded by longer grammatical forms” (Haspelmath 2019: 315).

# Previous studies on preposition-drop in Russian varieties

- Preposition-drop is reported to be typical of speakers of **very different** languages of Russia:
  - Daghestanian Russian (Daniel & Dobrushina, 2010: 77–80; **Panova & Philippova 2021**), Russian spoken in Mordovia (Pussinen, 2010: 119; Shagal, 2016: 370–372), Chuvash Russian (Baida 2018), Komi Russian (Baklanova, 2014: 124; Boronnikova, 2014: 119), Kalmyk Russian (Akimenko, 2014: 22), Tatar Russian (Boronnikova, 2014: 117), Shor Russian (Rezanova & Dybo, 2019: 201, 206); Kamchatkan Russian (Braslavets 1968: 319).  
→ At the same time, all these indigenous languages are **languages without prepositions**.
- Preposition-drop is attested in some monolinguals' Russian dialects (Myznikova 2014): probably, this is also a result of language contact.



## 2. Background information on Southern Tungusic and Northern Samoyedic



# Southern Tungusic and Northern Samoyedic in contact with Russian

- **Languages:**
  - Nanai and Ulcha - Southern Tungusic, the Amur region (Khabarovsk Krai).
  - Forest Enets, Tundra Enets, and Nganasan - Northern Samoyedic, the Taimyr peninsula
  - All these languages are endangered, Nganasan, Ulcha and both Enets especially are critically endangered.
- **Speakers:**
  - of older generation
  - (almost) all also speak Russian: some speakers acquired Russian at school age as a L2; some others acquired it at preschool age as their (near-)native language;
  - nowadays, Russian is used by bilingual speakers more actively than the IL.
- **Contact with Russian:**
  - both areas under discussion belong to those of the late Russian colonization.
  - the Taimyr peninsula: starting from the XVII–XVIII cent
  - the Amur region area inhabited by the Nanai and Ulcha: the middle of the XIX cent.
  - both in Taimyr and in the Amur region, intense contact between Russians and the indigenous peoples in question did not begin until the Soviet regime.

# Phonological and phonotactic restrictions

- Both in Southern Tungusic and in Northern Samoyedic, consonant clusters cannot occur word-initially

Avrorin (1959: 54–56) for Nanai, (Khanina & Shluinsky, in press) for Tundra and Forest Enets, Wagner-Nagy (2019: 37) for Nganasan

- Both in Southern Tungusic and in Northern Samoyedic, there are no exact correlates to the Russian /v/ and /f/, which are phonetic realization of the preposition ‘in’

Avrorin 1959: 32; Oskolskaya 2020: 311 on Nanai; Sunik 1985: 28 on Ulcha, Terešjenko 1966: 441 on Forest Enets; Khanina & Shluinsky, in press on Forest and Tundra Enets; Terešjenko 1979: 26; Wagner-Nagy 2019: 35–36 on Nganasan

# Cases and postpositions, no prepositions

	Nanai	Ulcha	Enets	Nganasan
Goal		postp/LAT		LAT-DAT/postp
Location		ESS-DAT/postp		LOC/postp
Source	EL/postp	INS/postp		ABL/postp
Trajectory	LOC/postp	PROL/postp		PROL/postp
Instrument		INS		LOC
Comitative			postp/LOC	postp/COM
Time ('last year', 'at 6 am')		ESS-DAT	LOC	LOC/postp
Recipient				LAT-DAT

### 3. 'In'-drop: statistical analysis

# Method

- Logistic regression model with mixed effects
  - (Gries, 2013: 333–336; Levshina, 2015: 192–196),
  - cf. a similar analysis for Daghestanian Russian in (Panova, Philippova 2021)
- R: function `glmer()`, package `lme4` (Bates et al., 2015)
- function `drop1()` to drop irrelevant predictors, basing on Akaike Information Criterion (AIC)
  - Gries, 2013: 266–267; Levshina, 2015: 266–267.

# Predictors

- **Phonetic context** ~ phonetic influence hypothesis
  - consonant-initial: <v> *tundr-u* 'to the tundra'  
[in tundra-ACC]
  - vowel-initial: <v> *Avam-sk-uju tundr-u* 'to the Avam tundra'  
[in Avam-ATTR-SG.F.ACC tundra-ACC]
- Expectations:
  - *v* 'in' will drop more easily before consonants than before vowels, since in this case consonant clusters appear, which breaks the rules of Tungusic and Samoyedic languages.

# Predictors

- **Case** ~ morphosyntactic influence hypothesis & non-contact hypothesis
    - prepositional - expresses Location: *v koridor-e* 'in the hall'  
[in corridor-PREP]
    - accusative - expresses Goal: *v koridor* 'into the hall'  
[in corridor.ACC]
  - Expectations:
    - only in Tungusic Russian, *v* 'in' will drop more easily in PPs with the prepositional case (which correspond to case forms in Tungusic) than in PPs with the accusative (which correspond to postpositional phrases in Tungusic)
- or
- in both varieties *v* 'in' will drop more easily in PPs with the prepositional case, because this case cannot be used without prepositions, so the preposition is redundant

# Predictors

- **Prototypical location** ~ non-contact hypothesis
  - “place names” — typical localities (<*v*> *poselke* ‘in the village’) and topographical objects (<*v*> *tundre* ‘in tundra’), toponyms, buildings and their parts (<*v*> *čume* ‘in the tent’), institutions (*v kolchoze* ‘at the collective farm’), spatial nouns (<*v*> *seredine* ‘in the middle’), and abstract spacial nominations (<*v*> *rajone* ‘in the region’).
  - “time names” (<*v*> *aprele* ‘in April’)
  - non-prototypical locations — all other uses.
- Expectations:
  - *v* ‘in’ will drop more easily in PPs referring to place and time names, in which the preposition is less informative & which tend to be expressed in a more reduced way cross-linguistically.

# Predictors

<b>Variable</b>	<b>Value</b>	<b>Tungusic Russian</b>	<b>Samoyedic Russian</b>
phonetic_context	consonant	54.40% (235    197)	23.7% (114    367)
	vowel	33.70% (31    61)	16.67% (12    60)
case	prep	49.84% (158    159)	23.64% (78    252)
	acc	52.17% (108    99)	21.52% (48    175)
prototypical_location	place	56.12% (165    129)	22.77% (69    234)
	time	60.20% (59    39)	35.71% (40    72)
	other	31.82% (42    90)	12.32% (17    121)

# Logistic regression: Tungusic Russian

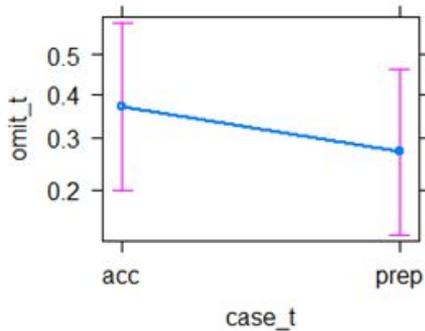
	Estimate	Std. Error	z value	Pr(> z )	Significance level
(Intercept)	-2.2427	0.5526	-4.059	4.93e-05	***
phonetic_context (consonant)	0.9865	0.2983	3.307	0.000944	***
case (prep)	-0.4664	0.2437	-1.914	0.055647	.
prototypical_location (time)	1.2001	0.3636	3.301	0.000965	***
prototypical_location (place)	1.2001	0.2872	4.179	2.93e-05	***

# Logistic regression: Samoyedic Russian

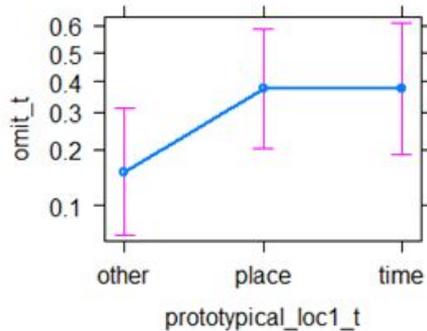
	Estimate	Std. Error	z value	Pr(> z )	Significance level
(Intercept)	-1.6279	0.5052	-3.222	0.001271	**
prototypical_location (time)	1.1719	0.3539	3.312	0.000928	***
prototypical_location (place)	0.5302	0.3165	1.675	0.093886	.

# Effect plots: Tungusic Russian, Samoyedic Russian

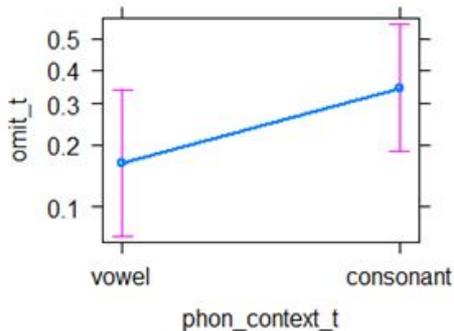
case\_t effect plot



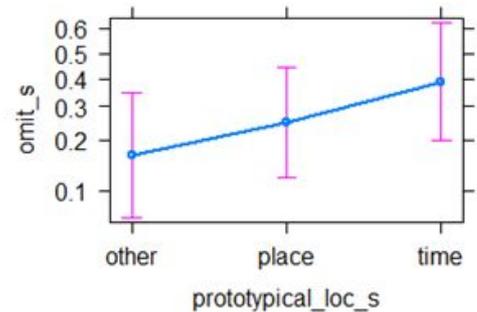
prototypical\_loc1\_t effect plot



phon\_context\_t effect plot

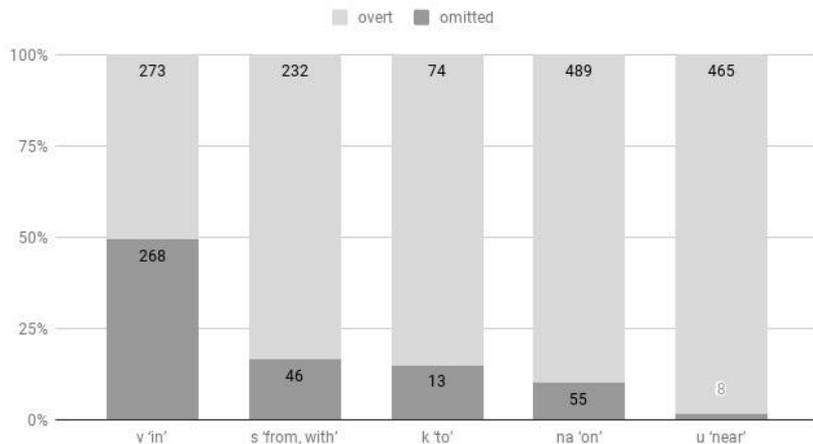


prototypical\_loc\_s effect plot

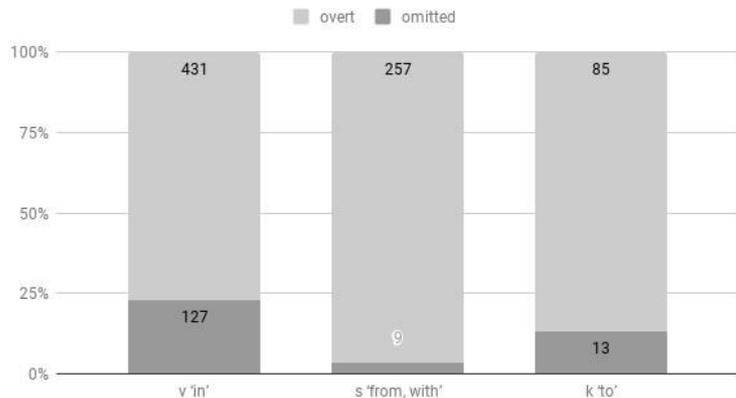


## 4. Other prepositions

Tungusic sample



Samoyedic sample



## Tungusic Russian:

- v 'in', s 'from, with', k 'to', u 'near'

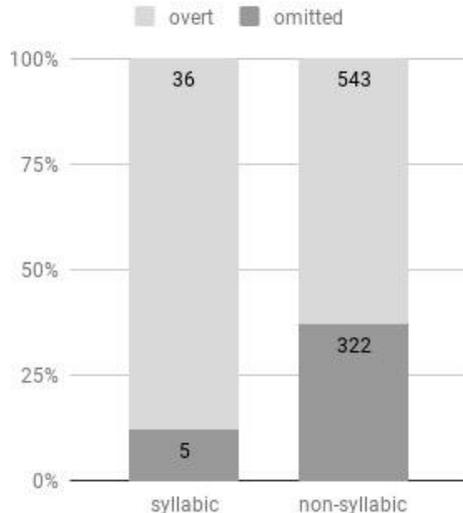
## Samoyedic Russian:

- v 'in', s 'from, with', k 'to' (NB only short ones)

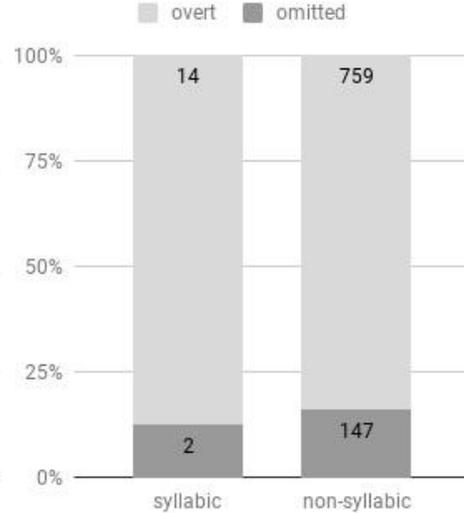
# Syllabic vs. non-syllabic variants of short prepositions

v 'in', k 'to', s 'from, with' ~ vo, ko, so

Tungusic sample



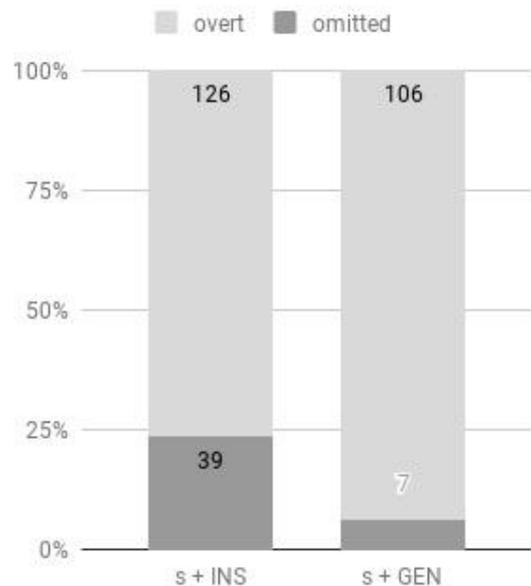
Samoyedic sample



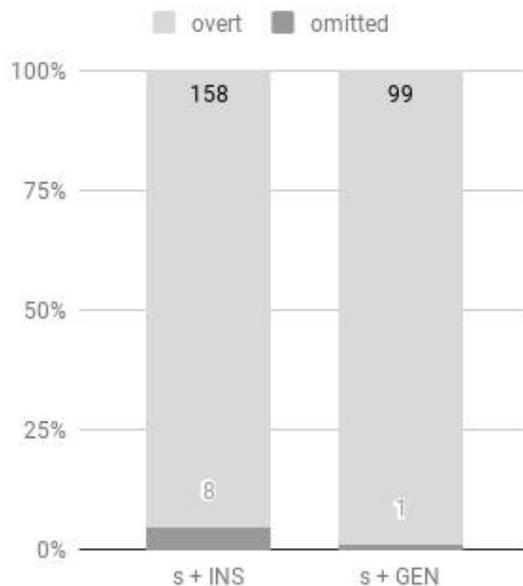
- Non-syllabic variants are omitted more often.
- This supports the phonetic hypothesis.
  - Tungusic sample: significant.
  - Samoyedic sample: non-significant.

# S 'with' vs. s 'from'

Tungusic sample



Samoyedic sample



- S + INS is omitted more than s + GEN
- The phonetics is not the only factor.
  - Tungusic sample: significant.
  - Samoyedic sample: non-significant.

# 5. Inter-speaker variation

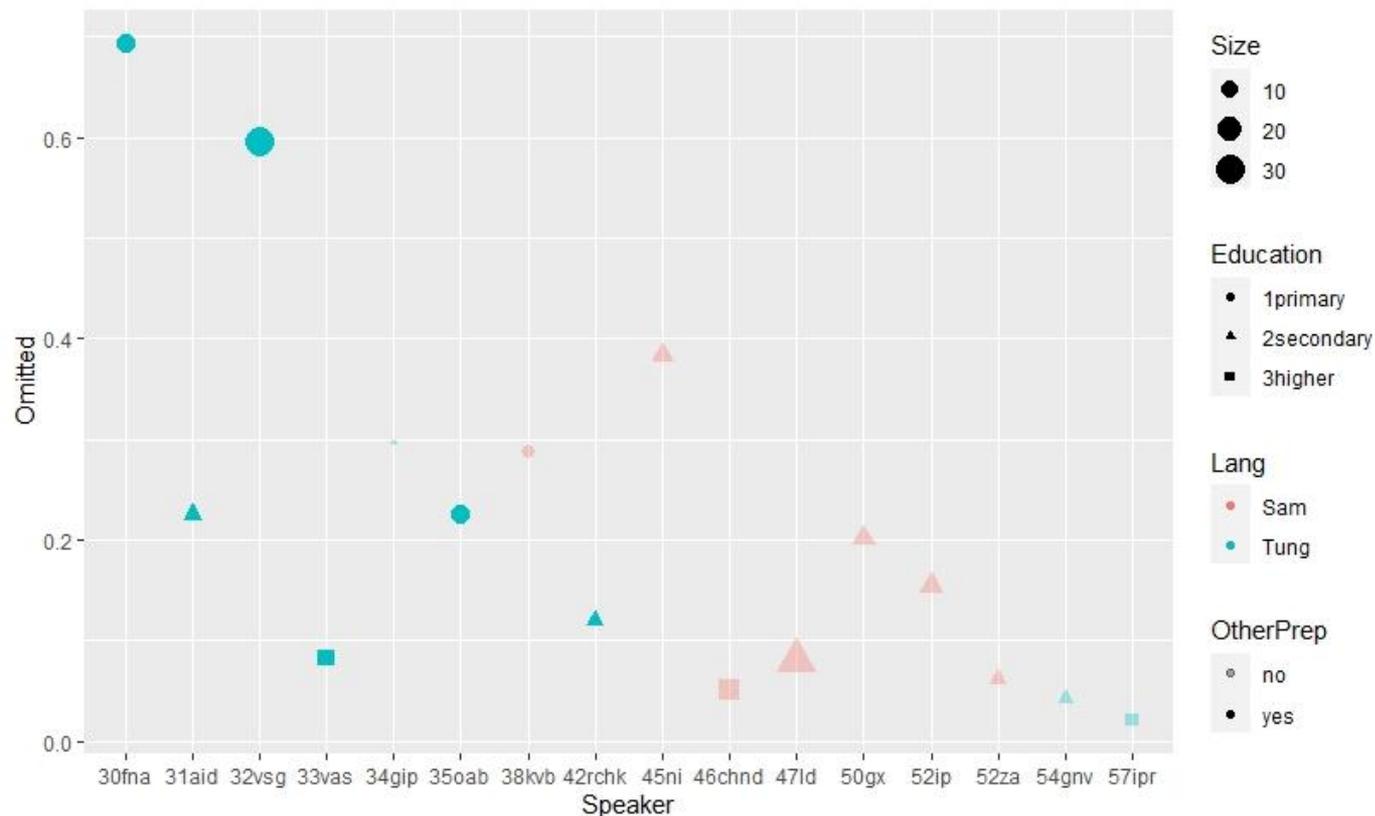
# Inter-speaker variation: sociolinguistic parameters

Year of birth: 1930-1957.

Level of education: primary, secondary, higher

L1: Tungusic (Nanai, Ulcha), Samoyedic (Tundra Enets, Forest Enets, Nganasan,  
+ some narrators speak Enets & Nganasan or Enets & Tundra Nenets)

# Inter-speaker variation: sociolinguistic parameters



- **L1:**
  - Samoyedic - low % of omissions
- **Education:**
  - no clear correlation
  - But all speakers with higher education - low % of omissions
- **Year of birth:**
  - no clear correlation
  - but starting from the 1950s all speakers have low % of omissions

# Inter-speaker variation: syntactic distinctiveness

- Annotation of contact-induced features implemented in the corpus
- Syntactic tags: disagreement, non-standard argument encoding etc.

(Khomchenkova et al. 2019)

*Za **dobryj pogoda** idi* (nio\_kvb) 'Go to get good weather'  
pp (*za xoroshej pogodoj* - NOM, m instead of INS, f)

*Nikakoj bog **ot** menja vyshe **netu*** (nio\_kvb) 'There is no god better than me.'  
neg (*nikakogo boga* - NOM instead of GEN),  
comp (*vyshe menja* - ot + GEN instead of GEN)

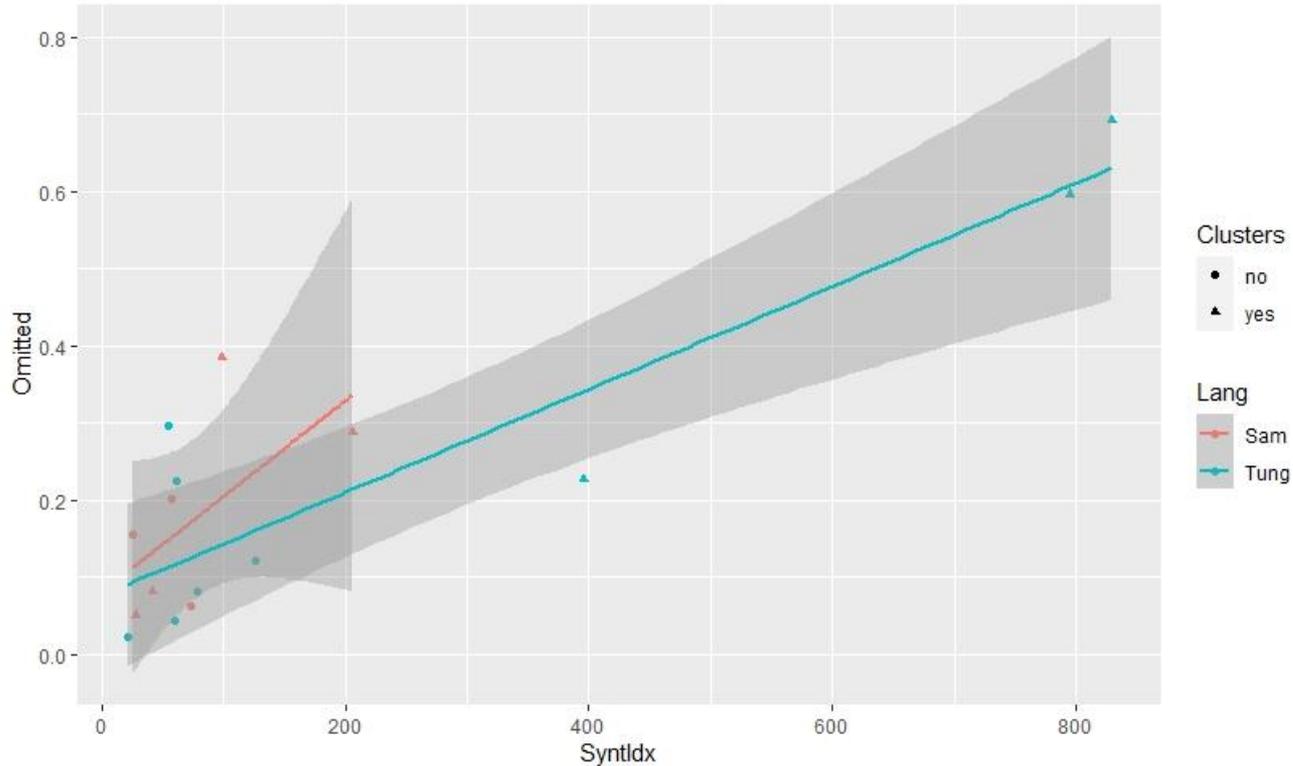
- Syntactic Index (for each speaker) = N of syntactic tags used / N of tokens  
→ If preposition drop is a morphosyntactic feature, then a correlation with Syntactic Index is expected. **YES (Tungusic sample)**

# Inter-speaker variation: cluster avoiding

- Some speakers tend to simplify consonant clusters:
  - *vtoroj* ~ *toroj* 'second'
  - *vkusnyj* ~ *kusnyj* 'tasty'
  - *starik* ~ *tarik* 'old man'

→ If preposition drop is a phonetic feature, then cluster-avoiding speakers are expected to drop prepositions more often. **NO**

# Inter-speaker variation: cluster-avoiding & syntactic distinctiveness



## 6. Examples

# Examples: preposition-drop

- (1) *Ja*                    *ži-l-a*                    *<v>*                    *Jakolevk-e.*  
I                            live-PST-SG.F                    in                    Yakolevka-PREP  
'I lived **in Yakolevka.**' (enh\_za)
- (2) *Oni*   *že*            *vot*            *<s>*            *mater'-ju*   *obš'a-l-i-s'*            *že.*  
they PTCL            here            with            mother-INS   talk-PST-PL-REFL   PTCL  
'They communicated **with their mother.**' (enh\_za)
- (3) *Vstreti-l-a*            *<na>*            *bazar-čik-e,*            *<v>*                    *Kondon-e.*  
meet-PST-SG.F            on            market-DIM-PREP            in                    Kondon-PREP  
'<I> met <this woman> **at the small market, in Kondon.**' (gld\_vsg)

# Examples: not preposition-drop

- (4) *I tam... osta-eš'-s'a svež-ij vozdux...*  
and there remain-NPST.2SG-REFL fresh-SG.M air  
'And there... you stay **in the open air.**' (enh\_ip)

- (8) *Ěto v ži= <za?> žimolost'-ju po-eha-l-i.*  
this-SG.N in for honeysuckle-INS PV-drive-PST-PL  
'We went to pick **honeysuckle.**' (ulc\_aid)

# Examples

Разные интересные примеры

В т.ч. На реинтерпретацию и