Olesya Khanina[1], Maria Ovsjannikova[2]
Institute of Linguistics, Moscow[1,2]/ Institute for Linguistic Studies, St. Petersburg[2]
olesya.khanina@gmail.com, masha.ovsjannikova@gmail.com

Enets non-finite clauses:
an intergenerational study of a seriously endangered language

1. Introduction

1.1. Enets: basic facts
- Samoyedic < Uralic;
- North of Central Siberia, Russia;
- two dialects: **Forest** (15–20 speakers), Tundra (10–15 speakers).

1.2. Data
- Corpus of glossed texts in Forest Enets (ca. 115 000 tokens / 21 hours):
  - archive recordings of 1970s–1990s ≈ speakers born in 1910s–1930s;
  - modern recordings made in 2005–2011 ≈ speakers born in 1940s–1960s;

1.3. Goal of the study
- The study aims at:
  - detecting the **changes in Enets syntax** that can be attributed to **language attrition** comparing the **frequency distributions** of non-finite forms in the texts of speakers with different years of birth (apparent-time study);
  - relate these changes to sociolinguistic history of the Enets community.

- Enets uses a wide array of non-finite forms in clause-combining: clausal complementation, noun modification, adverbial clauses.
- The use of non-finite forms can be expected to change under attrition due to:
  - language-internal processes of simplification of grammatical system and language use
  - influence of dominant language, Russian, which has few non-finite forms and their use is strongly associated with formal registers

1.4. Outline of the talk
- Recent sociolinguistic history of the Enets community
- Frequency and diversity of non-finite forms
- Case study: the patterns of clausal complementation

2. Sociolinguistic history of the Enets community

| The generation born in the 1910–1930s	vs.	The generation born in the 1945–1960s
- learnt Russian as adults - at short literacy courses organized in the tundra (*likbez*), - occasionally at the School of kolkhoz workers in the local town Dudinka		learnt Russian in schools at the age of 7-8 years old

- used Enets as the main language of communication with everyone in all domains, except for:
  - communication with the Russian newcomers, or other newcomers to the area;
  - in the new civilizational domains (in shops, hospitals, at post, at administration office, etc.)

- after schooling, used Enets as the main language of communication only
  - with representatives of the older generation;
  - while practicing traditional activities in the tundra
- Russian used in all educational contexts, everywhere in the villages/town, in mixed marriages, but also among Enets siblings and in Enets couples, i.e. with the non-Enets and the Enets of the same generation and younger

- By the end of the 1990s – beginning of the 2000s:
  - all Enets speakers stopped practicing traditional activities and relocated to villages (by age, by illness, by general collapse of the state-supported reindeer husbandry);
  - all representatives of the older generation (born 1910s–1930s) passed away.
- Thus, at the time when the modern recordings (2005–2016)were made the speakers have not used Enets for ca. 15 years.

3. Non-finite forms: frequency and diversity

3.1. Ratio of non-finite forms to all verb forms

- The most basic parameter is the ratio of non-finite forms to all forms in the texts by the same speaker.

Figure 1. Linear regression: Ratio of non-finite forms as a function of speaker's YoB and type of recording, with interaction term and overall number on verb forms in the speakers' texts as weights (reflected in the size of points)
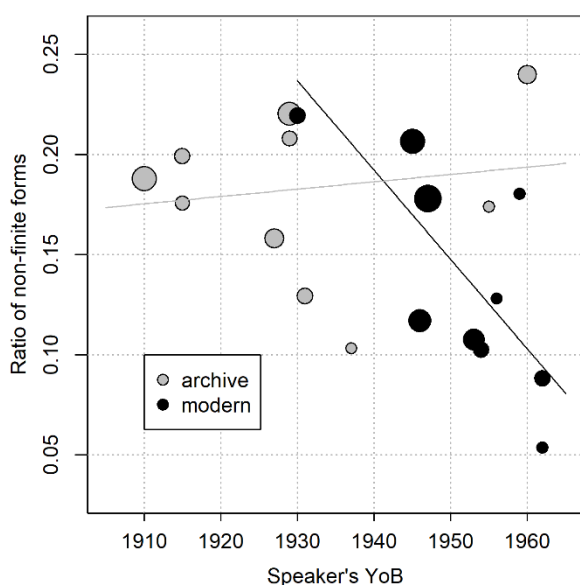

- Speakers from the **archive recordings**: no significant change of the ratio of non-finite forms, slight positive slope of the regression line is due to an unusual speaker born in 1960.

- Speakers from **modern recordings**: strong negative correlation between the ratio of non-finite forms and a speaker's YoB, i.e. younger speakers tend to use fewer non-finite forms than older speakers.

- For the modern recordings, the **major decrease** in the ratio of non-finite forms is observed for **speakers born after 1950**, the speakers born in the 1940s group together with speakers from the archive recordings.

2

3.2. Diversity of non-finite forms

- Language attrition often manifests itself in the loss of rare forms and levelling of fine-grained semantic distinctions, cf. [Campbell, Muntzel 1989; Dal Negro 2004] among others.
- The frequency distribution of Enets non-finite forms by speaker is shown in Table 1 — it is difficult to compare speakers directly due to large differences in the overall numbers of non-finite forms.
- To compare the speakers in terms of the inventory of non-finite forms and the evenness of their distribution we used the Shannon index of diversity[1]: higher Shannon index corresponds to greater diversity, see Figure 2.

Figure 2. Diversity of non-finite forms in the speakers' texts: the Shannon index
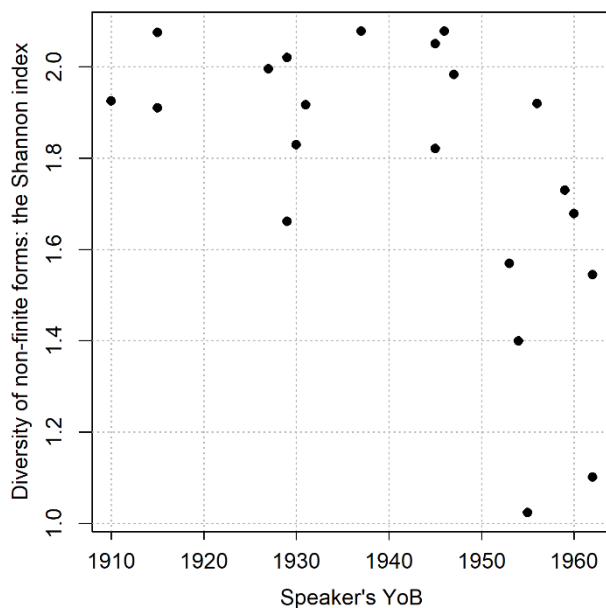
- In the texts by the **speakers born after 1950** the diversity of non-finite forms is lower than in the texts of the older speakers.

4. A case study: changes in the use of complement clauses

- For fine-grained distinctions, it is often difficult to establish trends and assess them statistically due to low number of occurrences for individual speakers.
- Looking for monotonous change in the occurrence of particular functions / constructions.

4.1. Complement clauses headed by nominalizations

- Nominalization in *-a* in core (1)–(2) and peripheral (3) syntactic positions typically occupied by nominal NPs:

(1) *kaʃi-tʃiʔ*   *kunin*   *dʲiri-e-xaru*   *dʲɔxara-xiʔ*
     man-OBL.PL.3DU   where/when.LOC.SG   live(ipfv)-NMLZ1-EVEN   not_know(ipfv)-3DU.S
     'They even don't know where the other one lives'.

---

[1] Calculated using the function entropy.ChaoShen() of the package entropy [Hausser, Strimmer 2014] for R [R Core Team 2017].

Table 1. Non-finite forms in Forest Enets: distribution by speaker

Speaker	CVB	PTCP.SIM	CVB.COND	PTCP.ANT	NMLZ	NMLZ+ABL	CVB.SIM	PTCP.ANT.PASS	CVB.IRR	SUP	PTCP.POST	PTCP.ANT.NEG	JUSS.NEG	CVB.ANT	N
NSP1910	0.25	0.19	0.21	0.15	0.09	0.05	<0.01	0.02	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	762
ASP1912	0.20	0.18	0.15	0.26	0.06	0.03	0.02	0.02	0	0	0.02	0.05	0.02	0	65
EDB1915	0.37	0.12	0.15	0.13	0.12	0.05	0.03	0.01	0.01	0.02	0	0	0	0	95
SPB1927	0.29	0.18	0.15	0.16	0.10	0.06	0.01	0.03	0.02	0.01	0.03	0.01	0	0	200
MNS1929	0.46	0.17	0.15	0.12	0.01	0.02	0.04	0.01	0.02	0	0	0	0	0	94
VNB1929	0.22	0.23	0.19	0.11	0.09	0.06	0.02	0.03	0.01	0.01	0.01	<0.01	0	<0.01	684
ND1930	0.39	0.22	0.13	0.06	0.04	0.06	0.03	0.02	0.02	0.04	0	0	0	0	120
MNB1931	0.36	0.18	0.18	0.06	0.12	0.01	0.03	0	0	0.01	0	0	0	0	67
NPCH1937	0.30	0.20	0.10	0.10	0.20	0	0	0	0.10	0.10	0	0	0	0	16
AP1945	0.29	0.08	0.05	0.07	0.12	0.13	0.17	0.02	0.02	0.04	0	0	0	0	95
NI1945	0.39	0.14	0.18	0.08	0.08	0.04	0.06	0.02	0.01	<0.01	0	<0.01	<0.01	0	838
NK1946	0.29	0.14	0.17	0.12	0.09	0.06	0.04	0.03	0.03	<0.01	<0.01	<0.01	0	0	294
LD1947	0.25	0.25	0.13	0.14	0.06	0.07	0.03	0.02	0.02	0.01	0.01	0	0	0	1363
AS1953	0.36	0.27	0.19	0.06	0.10	0	0.01	0.01	0	0	0.01	0	0	0	195
SA1954	0.50	0.25	0.06	0.02	0.15	0	0	0	0	0.02	0	0	0	0	48
EIB1955	0.63	0.06	0.28	0.03	0	0	0	0	0	0	0	0	0	0	32
GA1956	0.10	0.35	0.25	0.15	0.05	0	0	0.05	0.05	0	0	0	0	0	20
II1959	0.23	0.10	0.23	0.10	0.10	0	0.26	0	0	0	0	0	0	0	31
NNB1960	0.38	0.26	0.09	0.09	0.08	<0.01	0.10	0	0	<0.01	0	<0.01	0	0	231
AM1962	0.30	0	0.60	0	0.10	0	0	0	0	0	0	0	0	0	9
ES1962	0.44	0.27	0.20	0	0.02	0	0.02	0	0	0.02	0.02	0	0	0	45

(2) *tɛxɛ* *kudaxaa* *nʲi* *ŋa-ʔ,*
there(loc) for_a_long_time NEG.3SG.S exist(ipfv)-CONN
mu-za, **kaʔuʔ-a-za** *ɔzi*
PLC-NOM.SG.3SG **go_down(ipfv)-NMLZ1-NOM.SG.3SG** be_visible(ipfv).3SG.S
'It is not far away, the place where it flows in is seen'.

(3) *tʃike-d* **ɔ-ma-d** *dʲekoon*
this-OBL.SG.2SG **eat(pfv)-NMLZ1-OBL.SG.2SG** instead_of
kirba-zo-d *mu-ʔ* *anʲi*
bread-DEST.SG-OBL.SG.2SG take(pfv)-2SG.S.IMP and
'Instead of drinking this, buy bread!'

- The most frequent context of this type, as a complement of *kɔma* 'want' (4), is treated as a separate category in the counts:

(4) *ɔza* **oo-ma-d** *nʲi-ʔ* *kɔma-ʔ*
meat **eat(ipfv)-NMLZ1-DAT.SG** NEG-3PL.S want(ipfv)-CONN
'They don't want to eat meat'.

- Table 2 shows the frequency of uses in complement clauses, with the verb *kɔma* 'want' and all the other verbs, against the overall number of occurrences of NMLZ. (Other occurrences of NMLZ include mainly uses in various temporal adverbial clauses.)

Table 2. Nominalizations in *-a* in embedded clauses

	Verb *kɔma* 'want'		All the other complement clauses		Overall number of NMLZ occurrences
	N of occurrences	Ratio	N of occurrences	Ratio	
NSP1910	21	0.3	28	0.4	71
EDB1915	3	0.3	6	0.5	11
SPB1927	6	0.3	4	0.2	19
VNB1929	6	0.1	30	0.5	63
AP1945	4	0.4	3	0.3	11
NI1945	17	0.3	3	<0.1	62
NK1946	24	0.9	2	<0.1	27
LD1947	35	0.5	9	0.1	74
AS1953	16	0.8	2	0.1	19
NNB1960	7	0.4	7	0.4	18

- An increase in the frequency of uses with the verb *kɔma* 'want' and a decrease in the frequency of uses in the other types of complement clauses (Cochran-Armitage trend test, $p < 0.001$ in both cases).
- The decrease in the use of nominalization in all embedded clauses (except for those with 'want') can be at least partly attributed to the influence of Russian, where nominalizations are not used as a productive way of clause-combining.

4.2. Clausal complements of the verbs tɛne *'know'* and dʲɔxara *'not know'*

- Non-finite (1), (5) and finite (7)–(8) encoding of complement clauses:

(5) *ɔbu-xuru* *page* **sɔzuru-ʃ** *dʲɔxara-zʔ,*
what-EVEN outerwear **sew(ipfv)-CVB** not_know(ipfv)-1SG.S
nɔzu-dʲ *tɛne-zʔ*
scrape(ipfv)-CVB know(ipfv)-1SG.S
'I cannot sew any clothes, but I can scrape'.

(6) *ko-koz*   **tɛne-ʐa**,   *sɛn*   *pɔ*   **dʲiri-da-d**
     where-ABL.SG   know(ipfv)-3SG.SOsg   how_much   year   live(ipfv)-FUT-2SG.S
     'From where does it [a cuckoo bird] know, how many years you will live?'

Table 3. Distribution of non-finite and finite clausal complements
of verbs 'know' and 'not know' by speaker

	Non-finite	Finite
NSP1910	16	2
ASP1915	2	
EDB1915	3	
SPB1927	1	
MNS1929	6	1
VNB1929	5	
ND1930	1	
MNB1931	5	
NPCH1937	1	1
< 1940	**40**	**4**
AP1945	2	1
NI1945	4	3
NK1946	4	4
LD1947	11	12
AS1953	10	7
EIB1955		1
GA1956	1	
NNB1960	1	4
> 1940	**33**	**32**

- The finite strategy tends to be used more frequently in the texts by the speakers born after 1940s.

- The spread of finite constructions can be attributed to the interference from Russian.

4.3. Complement clauses headed by the -ʃ converb

- The -ʃ converb is used to mark the head of a complement clause for the majority of aspectual and modal complement-taking verbs, e. g. *pɛ* 'start' (7), *piris* 'can'.

(7) *tɔn*   **tɔda-go-ʃ**   *tɔz*   *pɛɛ-b*   *anʲi*
    now   **climb(pfv)-DUR-CVB**   so   start(pfv)-1SG.SOsg   and
    'So I started going up then'.

- The most frequent of these verbs is *tara* 'necessary' (8). It has been counted separately.

(8) *no, mɛ-ko-nʲʔ*   **kanʲe-ʃ**   *tara*
    well   chum-DAT.SG-OBL.SG.1DU   **leave(pfv)-CVB**   necessary(ipfv).3SG.S
   'Well, we had to go home'.

Table 4. Complement clauses with the -ʃ converb (*tara* 'necessary' vs. the other)

	tara 'necessary'		Other complement clauses		Overall number of CVB occurrences
	N of occurrences	Ratio	N of occurrences	Ratio	
NSP1910	4	0.02	73	0.41	179
ASP1915		0	5	0.39	13
EDB1915	11	0.34	10	0.31	32
SPB1927	6	0.11	7	0.13	53
MNS1929	5	0.14	12	0.33	36
VNB1929	16	0.11	25	0.18	140
ND1930	12	0.27	8	0.18	45
MNB1931	5	0.22	9	0.39	23
AP1945	2	0.08	3	0.12	26
NI1945	47	0.16	35	0.12	288
NK1946	14	0.21	11	0.17	66
LD1947	48	0.16	61	0.2	305
AS1953	27	0.4	14	0.21	67
SA1954	4	0.17	3	0.13	23
EIB1955	5	0.36	1	0.07	14
NNB1960	4	0.06	18	0.25	73
ES1962		0	3	0.16	19

- The frequency of occurrences of the converb in the complements of *tara* 'necessary' does not follow any discernible trend, while the other types of complements with this converb become less frequent in the texts by younger speakers (Cochran-Armitage trend test, p < 0.001).
- In complement clauses, the -ʃ converb is not likely to be displaced under the influence of Russian because it can be perceived as analogous to the Russian infinitive.

4.4. Frequency of aspectual and modal matrix verbs

Figures 3–7. Frequencies of five complement-taking verbs in the texts by the speakers born before 1940 and after 1940 (the frequency is measured in occurrences per one hundred)

pɛ 'start' *piris* 'can' *dʲɔxara* 'not know'

kɔma 'want' *tara* 'necessary'
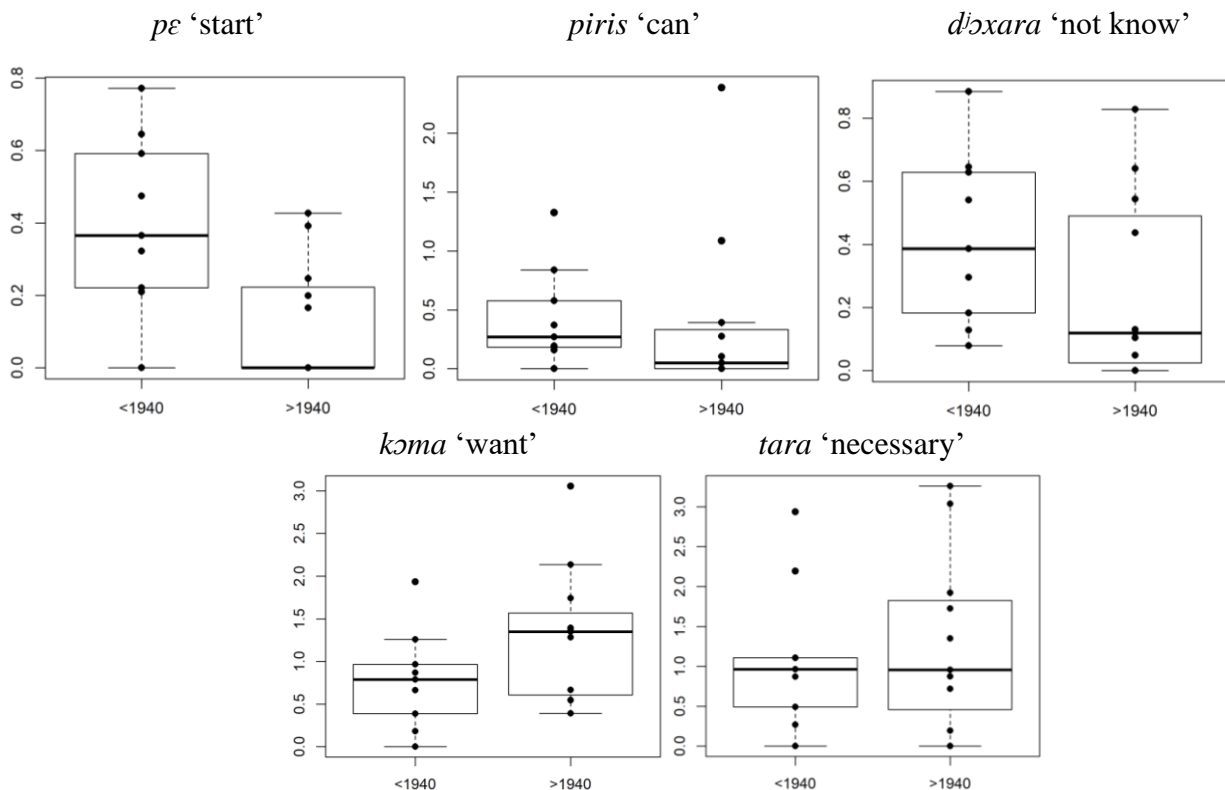
- In the texts by the speakers born before 1940, the verbs *pɛ* 'start', *piris* 'can' and *dʲɔxara* 'not know' are, on average, more frequent than in the texts by the younger speakers.
- The verbs *kɔma* 'want' and *tara* 'necessary' do not change their average frequency or become more frequent in the texts by the younger speakers.
- This observations suggest that younger speakers tend to use aspectual and modal complement-taking verbs less frequently than older speakers. Those verbs that are most frequent in the texts of older speakers have more chances to be preserved in the speech of younger speakers.

5. Conclusions

- Sociolinguistic factors and groups of speakers:
  - Speakers born before 1940: high frequency and diversity of non-finite forms and consistent use of non-finite strategies at the level of specific constructions.
  - Speakers born in the 1940s: high frequency and diversity of non-finite forms with interference phenomena at the level of specific constructions
    < acquisition of Enets in childhood, interference due to intense contact (and use) of Russian later in life.
  - Speakers born after 1950: decrease of frequency and diversity of non-finite forms, interference phenomena at the level of specific constructions
    < imperfect acquisition of Enets in childhood, interference due to intense use of Russian as adults.

- Language-external vs. language-internal causes of structural attrition:
  - The decrease in frequency of embedded clauses headed by nominalizations and the spread of finite strategies in the domain of clausal complementation is likely to result from the interference with Russian.
  - The decrease in frequency of complement clauses headed by "converb" can be attributed to syntactic simplification/reduction, viz. less frequent use of devices of aspectual and modal modification by younger speakers.
  - The decrease in the overall frequency and diversity of the non-finite forms can be attributed both to the interference from Russian and structural simplification/reduction.

References

Campbell L., Muntzel M. C. (1989). The structural consequences of language death // N. C. Dorian (ed.). Investigating obsolescence: Studies in language contraction and death. Cambridge: Cambridge University Press. P. 181–196.

Dal Negro S. (2004). Language contact and dying languages // Revue francaise de linguistique appliée. 2004/2. Vol. XI.

Hausser J., Strimmer K. (2014). entropy: Estimation of Entropy, Mutual Information and Related Quantities. R package version 1.2.1. https://CRAN.R-project.org/package=entropy.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.