Digital resources for Enets

A descriptive linguist's view

Olesya Khanina
Institute of Linguistics,
Russian Academy of Sciences, Moscow
olesya.khanina@gmail.com

Abstract: The paper describes digital resources available for Enets, an endangered Northern Samoyedic language: a multimedia dictionary and a digital corpus, both created by the author with her collaborators. Possible uses of these resources for linguistic research are discussed with the help of several examples. The dictionary is shown to seize limits of phonetic variation involving a mixture of two vowel phonemes, and the corpus is shown to provide a better description to grammatical structures involving ditransitive semantics. All examples illustrate the added value of the digital resources' use.

Keywords: corpus; dictionary; digital resources, Uralic; Enets; phonemic variation; ditransitives

1. Introduction

This paper is devoted to digital resources available today for Enets. Its aim is, first, to introduce the resources, and second, to give some examples of an effective use of these resources for descriptive linguistics. A particular emphasis is given to the added value of a description that is fed by the digital resources as compared to a description based on more traditional methodology and data. This paper has been written rather by a final-end user of computational applications than by a developer, which means it provides less description of technical details, but develops more on the ways the software can be used for research purposes.

Enets is a Northern Samoyedic, Uralic language spoken on the Tajmyr peninsula in the north of Central Siberia. It has two dialects, Forest and Tundra (FE and TE below) which are sometimes considered different languages (e.g., Siegl 2013 or Lewis et al. 2016). For the both dialects together, there are no more than 40 speakers left today, the language is not transmitted to children and is not used on everyday basis.

The digital resources for Enets existing today are two multimedia dictionaries, one for FE and one for TE, and a digital corpus, both open access for research purposes. First, I will present the dictionaries (section 2) and the digital corpus (section 3), and then I will introduce some puzzles

of Enets phonetics that could be solved with the help of the resources (section 4), followed by fragments of grammatical description of Enets fed by the digital corpus (section 5).

2. Multimedia Enets dictionaries

The two Enets dictionaries, one for FE, one for TE, with sound files and Praat annotations are currently under construction within a collaborative research project "Sound dictionaries of Uralic and Altaic languages of Russia" led by Anna Dybo and Julia Normanskaya at the Institute of Linguistics of Russian Academy of Sciences (Moscow). As a part of this project, specific software LINGVODOC was developed to create dictionaries for various Uralic and Altaic languages, and the software also allows making etymological connections for individual words within dictionaries from the same family. Individual dictionaries have some differences, as researchers have been given freedom in adapting the existing model to fit their specific data.

The Enets dictionaries contain Enets words in an IPA-based phonemic transcription designed by Andrey Shluinsky and the author, counterparts from the existing dictionaries (Sorokina & Bolina 2009) and (Helimski, manuscript),[1] a phonetic transcription in IPA, a translation into Russian, and at least one .wav sound file with several elicited pronunciations, often complemented by a Praat phoneme-by-phoneme annotation for at least one pronunciation, see Figure 1 and Figure 3. For many words, there are several audio files from different speakers. Some words feature "paradigm entries" where word forms are presented with phonemic transcriptions, phonetic transcriptions, translations into Russian, sound files, and eventual Praat annotations, see Figure 2 and Figure 4. Some words have etymological connections to words in other Uralic dictionaries created within the same software; these connections have been added by Julia Normanskaya. As can be seen from the given description, the Enets dictionaries are not yet dictionaries in strict terms of lexicography, as no syntactic information is now provided (parts of speech, case of arguments, etc.), and so for the moment, the label "dictionaries" can be used with this reservation only.

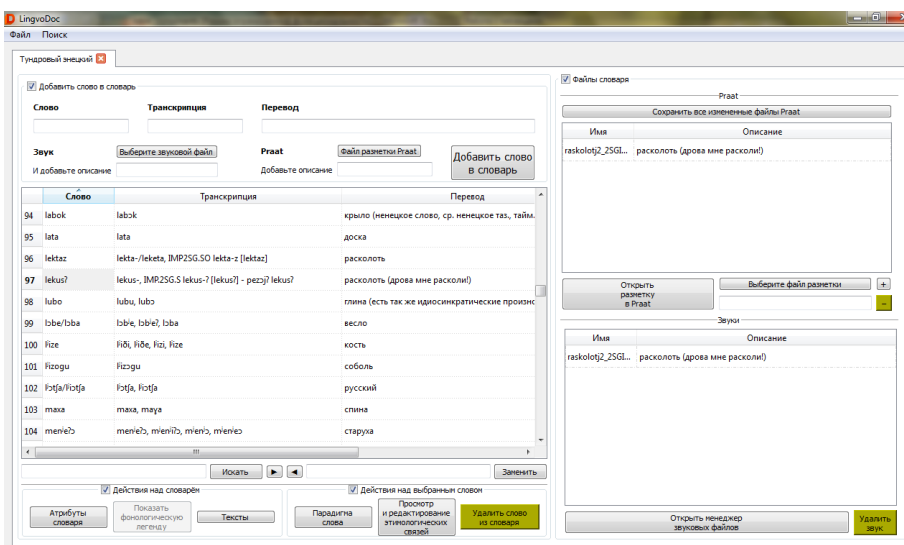[1] At the moment the counterparts are given in the FE dictionary only.

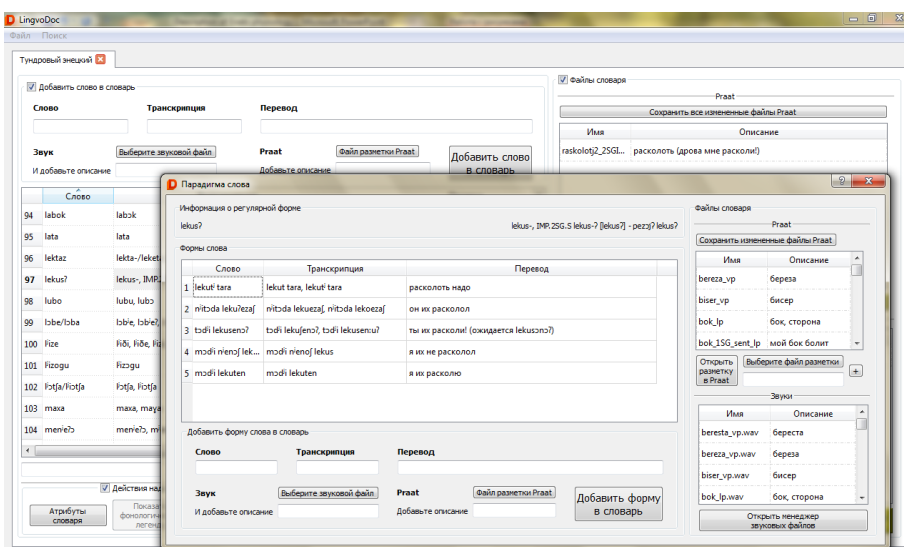**Figure 1:** Desktop version of the TE digital dictionary: word entries

Figure 2: Desktop version of the TE digital dictionary: a paradigmatic entry for verb *lekus* 'break (pfv)'

Figure 3: Web version of the TE digital dictionary: word entries

Figure 4: Web version of the TE digital dictionary: a paradigmatic entry for verb *lekus* 'break (pfv)'

All audio data for the Enets dictionaries have been collected in the field by careful elicitation with special attention to the quality of the recordings; both isolated pronunciations of target words as well as pronunciations of the same words in a phrasal context were recorded; Russian translation equivalents were used as stimuli for the elicitation. Before the sound files

could be incorporated into the dictionary, they had to be heavily edited so that the dictionary audio files featured only the target Enets words and occasionally their Russian translations, i.e., all other sound material had to be removed. The FE audio data were collected in 2008, 2015, 2016 in the village of Potapovo and in the town of Dudinka mainly by Andrey Shluinsky (Institute of Linguistics, RAS), but also by Maria Ovsjannikova (Institute for Linguistic research, RAS, in 2015, 2016) and the author (in 2008). The TE data were collected in 2008 in the village of Vorontsovo by Andrey Shluinsky and the author. The audio data for the both dialects were edited and entered into the LINGVODOC mainly by the author with initial help for FE from Andrey Shluinsky and Semen Sheshenin (Institute of Linguistics, RAS).

Figures 1 and 2 show a desktop version of the TE dictionary, here options for adding new data can be seen; Figures 3 and 4 show a web version of the TE dictionary, which is designed more for the display of the data, and a different logging in is required for editing the data (web version also has an English interface).[2] Note also that the dictionaries are being constantly expanded and updated, so what one can see in the Internet is very much a work-in-progress version.

At the moment, the TE dictionary has 256 lexemes with more than 1600 word forms and examples; the words feature pronunciations from two to four speakers. The FE dictionary has now 480 lexemes with more than 3300 word forms and examples; the words feature pronunciations from one to four speakers. For FE there are 1500 more lexemes collected that will be edited and entered into the dictionary later; a fieldtrip to TE is planned for having more TE lexemes recorded for the dictionary. The numbers of lexemes for the current Enets dictionaries are tiny in comparison to what is expected from a proper dictionary, so they cannot yet be used for any lexical or semantic studies. However, the number of word forms, all with phonetic transcriptions, is more impressive and already suffice for a phonetic, a phonological, or a morphological study. Section 4 will present some examples of phonetic research that was done on the dictionary data.

3. The digital corpus of Enets

The digital corpus consists of translated and glossed texts synchronized with audio (and for one third of the collection also synchronized with video). The transcription and translation were done in ELAN by Andrey

---

[2] http://lingvodoc.ispras.ru/

Shluinsky, Maria Ovsjannikova, Natalya Stoynova (Institute for Russian languages, RAS), Sergey Trubetskoy (Novosibirsk State University), and the author, and glossing was performed in Toolbox by Andrey Shluinsky and the author.[3] The Toolbox annotations can then be imported back into ELAN, and so ELAN files with transcription, translation, and glossing, all aligned with the sound, will be the final outlook of the corpus. Since the corpus is still being edited, it is available from its authors upon personal request, and in the nearest future it will be uploaded to the Internet for the ease of distribution and access.[4] The texts of the corpus are given in the same IPA-based phonemic transcription as was used for the dictionaries, with translations in Russian and English; before final uploading an Enets line in Cyrillic script will also be added. Therefore the corpus will be fully usable for speakers of English and Russian (given that the Enets community as well as many local researchers do not speak English, the latter is worth a particular mentioning). All texts of the corpus represent natural oral speech belonging to one of the following genres: everyday stories, traditional stories and tales, conversations, interviews, procedures/instructions, and songs. Due to the modern sociolinguistic setting of Enets, everyday stories make the majority of the collection; other genres are listed by the decrease of their share in the corpus.

The digital corpus has two subcorpora set by the two dialects, and two subcorpora set by the generation of speakers represented: current generation born in 1940s–1960s vs. generation of their late parents born in 1910s–1930s. Data for the latter were recorded in 1960s–1990s by linguists Kazimir Labanauskas, Eugen Helimski, Irina P. Sorokina, and Darja S. Bolina, by a musicologist Oksana E. Dobzhanskaja, and by a journalist of local radio Nina N. Bolina. In 2008–2011, these recordings were collected and digitized by Andrey Shluinsky and the author.[5]

Table 1: The size of the digital corpus of Enets

	Time in hours	Number of texts	Number of clauses	Number of tokens	Number of speakers recorded Current generation	Previous generations
Forest Enets	25	342	ca. 30 000	ca. 150 000	18	18
Tundra Enets	7.2	99	ca. 9 000	ca. 50 000	8	6
Total	**32.2**	**441**	**ca. 39 000**	**ca. 200 000**	**50**	

Table 2: Enets vowels. The /ɛ/ phoneme is found only in FE, otherwise the two dialects have identical sets of vowel phonemes and their allophones.

	Front	Central	Back
Close	i [i, ɨ]		u [u]
Close-mid	e [ə, e, ɛ, i, ɨ]		o [o, u, ɔ, ə]
Mid-open	ɛ [ɛ, æ, a]		ɔ [ɔ]
Open		a [a]	

Table 1 shows the size of the corpus: it is rather small in comparison to corpora of major world's languages, but quite big for a small indigenous language. Noteworthy, the same reservation has to be made here as to the label "corpus" for this resource on Enets, as was done above for the label "dictionary". The last decade has seen many corpora of small indigenous languages created by linguists all over the world, though none of them can be called "corpus" in the strict sense of the word used by corpus linguistics dealing with major languages, such as English, German, or Russian.

The Toolbox version of the corpus can easily be searched for any particular morpheme thanks to the in-built search option of Toolbox, the same is true for the ELAN version with its own in-built search, besides the latter also allows for regular expression searches.

4. A phonetic description fed by the digital Enets resources

Enets is a language with a very high phonemic variation: most phonemes have allophones, and in many cases the distribution between the allophones looks like a free one even after months of phonetic research, cf. Tables 2 and 3 (see also Siegl 2013, 82–110 for numerous examples).

Table 3: Enets consonants. The [tʃ] allophone of the /dʲ/ phoneme and the [s] allophone of the /z/ phoneme are found only in FE, otherwise the two dialects have identical sets of consonantal phonemes and their allophones.

	Bilabial	Dental/Alveolar	Palatalized coronals	Palatal	Velar	Glottal
Plosive	b [b], [bʲ], [p]	d [d]	dʲ [dʲ], [tʲ], [tʃ]		k [k], [kʲ], [x]	ʔ [ʔ, ∅, V̰]
	p [p], [pʲ]	t [t]	tʃ [tʃ], [tʲ]		g [g], [gʲ], [k]	
Nasal	m [m], [mʲ]	n [n]	nʲ [nʲ]		ŋ [ŋ], [ŋʲ]	
Trill		r [r], [rʲ]				
Fricative		z [z], [zʲ], [ð], [s]	ʃ [ʃ], [ʃʲ], [ç]		x [x], [xʲ]	
		s [s], [sʲ], [θ]				
Approximant				j [j]		
Lateral approximant		l [l]	lʲ [lʲ]			

Besides, many phonemes or their positional variants have zero realizations, e.g., glottal stops (1), final vowels (2), or second vowels in the sequence of two identical vowels (3).[6] In all these cases, no patterns regulating zero vs. non-zero realizations could be found again after months of phonetic research (i.e., targeted elicitation and its analysis).

(1) FE, TE /baʔa/ 'bed' [baʔa], [baa], [baː]

(2) FE /kɔdo/ 'sledge' [kɔdo], [kɔd], [kɔt]
     TE /nenogo/ 'mosquito' [nenɔgɔ], [nenɔg]

(3) FE /ʃee/ 'who' [ʃʲee], [ʃʲeː], [ʃʲe]
     TE /miiʔ/ 'what' [mʲi], [mʲiː], [mʲiiʔ]

A good illustration for this mess is the fact that, with an exception of the late Eugen Helimski (cf. Helimski, manuscript), none of the major publications on Enets has used consistent phonological spelling (see, e.g., Bolina 2012; 2014; 2015; Sorokina 2010; Sorokina & Bolina 2001; 2005; 2009; Labanauskas 1992; 2002; Siegl 2013), though it was only Florian Siegl (2013, 33) who explained his choice:

> "I have decided against both an abstract phonological transcription and normalization in order to preserve the encountered picture. At present normalization would be counter-productive; idiolectal variation in pronunciation, e.g., realization of glottal stops, alternating vowel length in identical forms, or the impossibility to identify a single underlying form which would be representative of

---

[6] If double vowels are analyzed as long vowels, then this can be reformulated: long vowels are often realized as short vowels.

> 'Forest Enets', do not justify any abstract normalization from the point of view of language documentation."

Indeed, I could sign under this quotation before I had the digital resources for Enets. There seemed to be no limits on the variation, at least no limits that could easily be identified, either by previous researchers or by myself. The first principles governing the variation started to emerge when Andrey Shluinsky and myself created the digital corpus of Enets (which was created before the multimedia dictionary). e.g.,, we noticed that there are dozens of words with [e] ∼ [i] ∼ [ɨ] alternation in non-first syllables and dozens of words with [i] ∼ [ɨ] variation in non-first syllables, but not a single word with only [e] in non-first syllable[7] (given that only words that occurred more than three times in the corpus were considered for the purpose). We could notice this since we had a technical "abstract phonological transcription" for each word in the Toolbox database, but allowed words to have "alternates" specified in the dictionary. Our findings could be reformulated as the following phonological generalization: all /e/ in non-first syllables can be realized as [e], [i], and [ɨ], while /i/ in non-first syllables can be realized as [i] or [ɨ] only.[8]

However, there were other numerous instances of variations that could not be explained with the data from the digital corpus only. Noteworthy, the digital corpus does not feature phonetic transcription, so we could study only the issues that we as corpus authors felt as probably phonological and thus requiring registering as alternates in the Toolbox database. So here came in the phonetic data from the digital dictionaries: there were not only several thousands of words with phonetic transcriptions, but also easily accessible audio data allowing easy re-listening with a subsequent change in phonemic transcription, if necessary. In this paper, I will discuss just one example of a phonetic puzzle solved with the help of the data and refer to (Khanina, Subm.) for more cases.

The earliest source on Enets, Castrén (1854/1969), mentioned only two back vowel phonemes, /ɔ/ and /u/. Modern recordings, as well as the recordings from the previous generation show variation [ɔ] ∼ [o] ∼ [u] for

---

[7] There are two FE homonymous morphemes, originating in one historic source, for which only [e] and no [i]/[ɨ] variants have been attested in a non-first syllable: these are regular allomorphs *-eʔ* of 1PL.S/SOsg and NOM.SG.1PL *-aʔ* (used with *i*-final stems); TE uses only *-aʔ* for all types of stems.

[8] Siegl (2013, 93–96) does not say anything about the [e] ∼ [i] variation, though there are words in the grammar that are spelled inconsistently both with *e* and *i*, e.g., *äsi* on pp. 235-236 or *äse* on p. 197 for 'father', *kańe-* on p. 197 or *kańi-* on p. 198, p. 200 for 'to go', *ote-* on p. 298 or *oti-* on p. 430 for 'to wait', etc.

some of the Castrén 1854's /ɔ/, but not for others, so there is some sound change going on in some words. Without reference to historical sources, the situation can be formulated this way (see also Table 2): there are words with /ɔ/ without any variation in the realization of the vowel, e.g., FE, TE /ɔdi/ 'young man', FE, TE /dʲɔtu/ 'goose', FE /ɔtʃik/ / TE /ɔptʃiko/ 'bad', there are words with /u/ without any variation in the realization of the vowel, e.g., FE, TE /buja/ 'blood', FE, TE /tʃuku/ 'whole', FE, TE /pu/ 'stone', and there are words with variation [ɔ] ∼ [o] ∼ [u], e.g., FE, TE /koba/ 'skin', FE, TE /to-/ 'come', FE, TE /kɔdo/ 'sledge'. So modern Enets, both FE and TE, has three back vowel phonemes, /ɔ/, /u/, and /o/: the same descriptive decision is taken in Tereščenko 1966; Susekov 1977; Sorokina & Bolina 2009; Sorokina 2010, Helimski, Ms. Though Siegl (2013, 95) explicitly doubts the existence of the three back vowel phonemes, he notes that some lexemes "were resistant and clearly showed /o/ where expected", while for most other words "the contrast between /u/ and /o/ is often neutralized", so de facto he also admits the existence of the three way opposition: "clear *o*" (my /ɔ/) vs. "clear *u*" (my /u/) vs. "neutralized contrast between *o* and *u*" (my /o/).

The descriptive question, then, is how one can know in which words the sound change /ɔ/ > /o/ has happened, and which words have opposed the change. In other words, when one hears an [ɔ], how can one know which phoneme it is, /ɔ/ or /o/?[9] Primarily, this is a bold descriptive problem, as each word has to be written down phonemically, and two or three instances are often not enough to learn the spelling of a word with [ɔ]. This problem could not be solved before the data from the dictionaries became available. Having closely studied the distribution of [ɔ], [o], and [u] in the latter, I managed to restrict the descriptive problem to the first syllable, formulating a set of straight rules for non-first syllables.

Indeed, the data suggest unambiguously that in non-first syllables Enets words contain mainly /o/, while /ɔ/ is possible only in the very limited contexts. First, /ɔ/ is possible after a glottal stop, and in this context there are dozens of instances of /ɔ/, e.g., TE /batoʔɔ/ 'back, tail', TE /baxoʔɔ/ 'old man, husband', FE /sokoʔɔte/, TE /sɔkoʔɔte/ 'sokuj, a traditional men's coat from fur', FE /sɔʔɔ-/ 'jump up (pfv)', and only a couple of instances of /o/ in FE, e.g., FE /seʔo/ 'seven', /ŋaʔo/ 'duck', /ezeʔo/ 'runner'; no /o/ after glottal stop is attested in TE. Second, /ɔ/

---

[9] A similar problem can be supposed for /o/ vs. /u/, but it does not arise in reality, as [ɔ] and [o] pronunciations for /o/ are much more common than [u] pronunciations. So two or three instances of the same word are usually enough to figure out whether it has /o/ or /u/.

is possible after another vowel, e.g., FE, TE /tɔɔ/ 'summer', TE /iriɔ/ 'moon', TE /sedeɔ/ 'past, former', FE /sɔɔko/ 'younger brother, youngster', /batoɔ/ 'tail, bottom', though the context is not very frequent in TE and very rare in FE (all three FE words of this type are listed here). After vowels, /o/ is possible only in very frequent affixal morphemes, e.g., FE, TE focus marker /-xoo/. Finally, in FE /ɔ/ is possible when it results from a recent irregular sound change /a/ > /ɔ/, with TE cognates having /a/ and sometimes with FE words having variation between /a/ and /ɔ/, e.g., FE /d'ɔxazi/ ∼ /d'ɔxɔzi/ 'female reindeer', in these words /o/ is equally possible, though much rarer, e.g.,/boo/ 'bad', cf. TE /bɔa/. Beyond these three contexts – after a glottal stop, after a vowel, or on the place of the recent /a/ – /ɔ/ is impossible in non-first syllables. Practically, it means that hearing the [o] allophone in a non-first syllable, one can immediately know that it belongs to the /o/ phoneme, unless one of the three conditions just discussed is met.

This discovery was not possible until hundreds of pronunciations with open back vowels /ɔ/ and /o/ were, first, spotted as such, and then studied in details.

5. A fragment of Enets grammatical description fed by the digital corpus

Enets has two ditransitive constructions, one of them features a Dative morpheme, as in (4), and the other features the so-called Destinative morpheme (or Benefactive in Siegl 2013), as in (5). The Destinative morpheme is cross-linguistically unusual: in transitive clauses, it marks the presence of a beneficiary on the direct object, while a possessive affix immediately after it expresses the recipient/beneficiary itself.[10]

(4)   FE
      prɔdaves nɛ-**d**   pɛɛ-nʲʔ   miʔɛ-zʔ   modʲ
      seller   woman-**DAT.SG** traditional_shoe-PL.1SG give(pfv)-1SG.S I
      'And so I gave my shoes to the saleswoman.'   (VNB9306_KAKZHI_052)[11]

---

[10] More rarely, a separate possessive noun phrase can be used to express a non-pronominal recipient/beneficiary. Note that the Destinative morpheme's uses are not restricted to transitive clauses, as examples (7)–(8) show, see also (Khanina & Shluinsky 2014) for more details.

[11] All examples are given with unique identifiers referring to the Enets corpus: the identifier features, first, speaker's or speakers' initials, then the date of recording in the format YYMMDD, then abbreviated name of the text, optionally followed by initials

(5) FE

sɔjza kɔru-**zo**-nʲʔ ta-ʔ

good knife-**DEST.SG**-OBL.SG.1SG give(pfv)-2SG.S.IMP

'Give me a good knife.' (AS_NI100713_RAZ_011)

The digital corpus provided data for an accurate description of the formal peculiarities of these constructions and their mutual distribution. Both issues could not be described without this resource, as is shown below with reference to a recent grammatical description (Siegl 2013) based mainly on elicitation.

The Destinative morpheme is -*zo*- for singular nouns and -*zi*- for plural nouns, for both Enets dialects. Already here a problem arises: Siegl (2013, 383) points out for FE that "when a NP is marked for BEN, overt number marking is absent". Is this indeed so? Our FE subcorpus contains 130 clauses with plural Destinative, as in (6), so we can learn from the corpus that number marking is definitely possible when a NP is marked for Destinative.

(6) FE

piinure-za buniki-**zi**-nʲʔ pɔzaru-nʲi-j-ʔ

be_frightful(ipfv)-PTCP.SIM dog-**DEST.PL**-PL.1DU harness(pfv)-CONJ-1DU.S/SOsg

'Let's harness frightful dogs-for-us (as reindeer)!' (NSP92_DVA_033)

Besides, Destinative marked on an adjunct is problematic for Siegl's (2013, 386) description: despite earlier accounts in the literature, he is unsure whether this construction really exists in modern FE, as he encountered only three examples of it in his elicitation sessions and no natural examples. Based on the digital corpus, a clear answer to it can be provided: there are 41 uses of adjunct Destinative in the FE subcorpus, as in (7), and adjunct Destinative is in fact twice more common than subject Destinative, as in (8), discussed by Siegl (2013) as definitely existing in FE.

---

of the speaker, if more than one speaker produced the recording, and finally the sentence number in the text.

The abbreviations in the glosses stand for: 1 = 1st person, 2 = 2nd person, 3 = 3rd person, ABL = Ablative, CONJ = Conjunctive, CVB = Converb, DAT = Dative, DEB = Debitive, DEST = Destinative, DU = Dual, FOC = Focal, IMP = Imperative, ipfv = imperfective, NOM = Nominative, OBL = Oblique, pfv = perfective, PL = Plural, PLC = Placeholder, PST = Past, PTCP.SIM = Simultaneous participle, S = Subject cross-reference, SG = Singular, SOsg = Subject-object cross-reference for singular object.

(7) FE
     tɔnneda mu-**zo**-naʔ   mɔzara-ʃ,
     then   PLC-**DEST.SG**-OBL.SG.1PL work(ipfv)-3SG.S.PST

     starʃij glavnij vetvratʃi-**zo**-naʔ   mɔzara-ʃ
     senior principal veterinarian-**DEST.SG**-OBL.SG.1PL work(ipfv)-3SG.S.PST
     'He worked then as this one of ours, he worked as the senior principal veterinarian.'
                                                    (SPB_NNB910131_INT_SPB_211)

(8) FE
     ɔbu-xoa   dʲa-xaz   entʃe-**do**-jʔ
     what-FOC land-ABL.SG person-**DEST.SG**-NOM.SG.1SG

     to-tʃu   anʲi
     come(pfv)-DEB.3SG.S and
     'A person for me will come from some land.'   (VNB9111_UROD_194)

Finally, Siegl (2013, 391–393) claims only two intransitive verbs to allow for their subjects to be marked by Destinative, i.e., he postulates a lexical restriction on this category. Again, the corpus data easily show that this stipulation is wrong: in the FE subcorpus there are five lexical intransitive verbs and a dozen of transitive verbs in Passive attested with Destinative subject.

The digital corpus features 393 clauses with the Destinative morpheme and 77 clauses with the Dative morpheme in the same function (i.e., nor locative uses of Dative, neither its argumental uses have been counted here). This set of clauses expresses both ditransitive and benefactive events, and the contexts look very similar, cf. (4) and (5).

Siegl (2013, 394–395) suggests one distributional pattern for the two ditransitive constructions: the ditransitive construction with Dative may be used when a transfer of possession is observed, while the ditransitive construction with Destinative is used when there is no transfer of possession. For instance, in (5) above, the speaker uses the Destinative construction to ask for a knife to cut a fish at the home of the addressee, so in this case no transfer of possession would indeed happen. So this observation holds true for some uses of the Destinative construction, but there are also clear counterexamples to it: cf. (9) where the fish would belong to the addressee, i.e., the transfer of possession definitely takes place.

(9) FE
     modʲ tɛxɛ   tɔrse kare-**zo**-d   tɔzu-ta-zʔ
     I   there(loc) such fish-**DEST.SG**-OBL.SG.2SG bring(pfv)-FUT-1SG.S
     'I will bring such fish for you here.'   (NI090719_ZOL_091)

Besides, this generalization does not account for the benefactive uses of both ditransitive constructions, illustrated in (6) for Destinative: these uses are numerous for the Destinative construction, and no transfer of possession can be postulated for any of them. However, a study of all corpus examples with their respective contexts suggests a pattern of the distribution covering all uses of the two seemingly homonymous constructions. It turns out that the Destinative construction introduces a new thematic referent for which there exists a recipient or a beneficiary. In contrast, the Dative construction highlights the known referent of a ditransitive/benefactive event. This known referent may be in the theme position, as in (4) or this known referent may be in the recipient/beneficiary position, as in (10).

(10) TE
      {I have relatives in Vorontsovo. [...] I have a brother and a sister. [...] I cannot go there
      now. [...]}
      mii-goa,   sɔjza mii-goa   mi-tʃi,   kaa-**xa**-nʲiʔ
      what-FOC good what-FOC give(pfv)-CVB relative-**DAT.SG**-OBL.SG.1SG
      '(At least) to send something good to my brother.' (ELSNNB970514ELS_INTW_192)

Although this description of the distribution is quite crude (see Khanina & Shluinsky submitted for a more fine-grained analysis), it shows well that factors in play could not be spotted without a representative corpus, preferably digital, for the ease of extracting all the necessary clauses and quick referral to their contexts.

6. Conclusion

The digital resources for Enets have helped to solve two puzzles connected to the allophones of some vowels phonemes in Section 4, but they could do even more. Khanina (2017) discusses the contexts for zero realizations of an intervocalic glottal stop and Khanina (submitted) – zero realizations of final vowels. The claims have been made in the literature, primarily in (Siegl 2013), that both final vowels and intervocal glottal stops have disappeared from modern FE, but the data from the FE multimedia dictionary show that this is not true, and allow for outlining the conditions of the zero phonetic realizations. Besides, optional consonant palatalization before front vowels can be studied in details, and even though at first no apparent rules governing the palatalization are visible, a study of frequencies of palatalizations for each consonant before each front vowel reveals the logic of this phonetic process.

The digital corpus was prepared earlier than the multimedia dictionaries, and so more descriptions realized with its help have already been published, see Khanina & Shluinsky (2016) for a thorough description of all uses of the Enets Perfect, Khanina (2016) for a description of a mutual distribution of various coordination strategies in Enets, Shluinsky (2010) for a first description of a FE complex mood not mentioned in any published sources, but proliferate in the digital corpus, or Khanina & Shluinsky (2015) for an attempt to describe rules for differential object marking realized by verbal cross-reference.

To sum up, digital corpora and dictionaries with audio data open wide research opportunities. First, with the help of their in-built search function and easy access to broader context, one can quickly obtain raw data to study patterns conditioned by discourse structure. Second, frequency statements become possible, and, more generally, a study of real language use and of actual concurring strategies for encoding seemingly "the same" kind of information becomes doable. Third, the data stored in digital dictionaries and corpora are significantly less dependent on current linguistic theories than data elicited to answer specific research questions, and so can be used again and again decades after the resources have been created to check hypothesis we cannot even imagine yet. Forth, such digital resources have crucial methodological advantages as compared to traditional resources, as they allow for an immediate verification by other researchers and tracking any scientific result back to the raw data; on the importance of this kind of verification for linguistics, see Bright (2007); Broeder et al. (2011); Chelliah (2001); Mithun (2014), and Mosel (2014).

Acknowledgements

References

Bolina, Zoja N. 2012. Èneckij kartinnyj slovar [Enets picture dictionary]. Dudinka: Tajmyrskij okružnoj centr narodnogo tvorčestva.

Bolina, Zoja N. 2014. Pesni rodnoj zemli: samodejatel'nyj èneckij xudožnik Ivan Silkin [The motherland songs: Ivan Silkin the amateur Enets painter]. Dudinka: Tajmyrskij okružnoj centr narodnogo tvorčestva.

Bolina, Zoja N. 2015. Ezzuuj. Sled narty [The sledge print]. Dudinka: Tajmyrskij okružnoj centr narodnogo tvorčestva.

Bright, William. 2007. Contextualizing a grammar. In T. Payne and D. Weber (eds.) Perspectives on grammar writing. Amsterdam & Philadelphia: John Benjamins. 11–17.

Broeder, D., H. Sloetjes, P. Trilsbeek, van Uytvanck D., M. Windhouwer and P. Wittenburg. 2011. Evolving challenges in archiving and data infrastructures. In G. L. J. Haig, N. Nau, S. Schnell and C. Wegener (eds.) Documenting endangered languages: Achievements and perspectives. Berlin & New York: De Gruyter. 33–54.

Castrén, Matthias Alexander. 1854/1969. Grammatik der samojedischen Sprachen. Hrsg. von Anton Schiefner = Nordische Reisen und Forschungen 7. St. Petersburg & Leipzig: Buchdruckerei der Keiserlichen Akademie der Wissenschaften.

Chelliah, Shobhana L. 2001. The role of text collection and elicitation in linguistic fieldwork. In P. Newman and M. Ratliff (eds.) Linguistic fieldwork. Cambridge: Cambridge University Press. 152–164.

Helimski, Eugen. Materialy k slovarju èneckogo jazyka [Materials for an Enets dictionary]. Ms.

Khanina, Olesya. 2016. Sočinitel'nye strategii eneckogo jazyka [Enets coordination strategies]. Uralo-Altaic Studies 21. 131–148.

Khanina, Olesya. 2017. Korpus i elektronnyj slovar' eneckogo jazyka: vozmozhnosti cifrovyh tehnologij [A corpus and an electronic dictionary for Enets: Possibilities open by digital technologies]. Uralo-Altaic Studies 3 (to appear).

Khanina, Olesya. submitted. Documenting a language with phonemic and phonetic variation: The case of Enet. Language Documentation and Conservation .

Khanina, Olesya and Andrey Shluinsky. 2014. A rare type of benefactive construction: Evidence from Enets. Linguistics 52. 1391–1431.

Khanina, Olesya and Andrey Shluinsky. 2015. Pr'amoj objekt v èneckom jazyke: objektnoe soglasovanie glagola [Direct object in Enets: verbal object cross-reference]. In L. E. A., Z. A. V. and K. M. B. (eds.) Tipologija morfosintaksičeskih parametrov: materialy mezhdunarodnoj konferencii "tipologija morfosintaksičeskih parametrov" [Typology of morphosyntactic parameters: Materials of the international conference "Typology of morphosyntactic parameters"]. Moscow: MPGU. 392–410.

Khanina, Olesya and Andrey Shluinsky. 2016. Eneckij perfekt: diskursivnye upotreblenija u evidencial'no-admirativnogo perfekta [Enets perfect: discourse uses of an evidential-admirative perfect]. Acta Linguistica Petropolitana 12. 425–474.

Khanina, Olesya and Andrey Shluinsky. submitted. Competing ditransitive constructions in Enets. Folia Linguistica .

Labanauskas, Kazimir I. 1992. Folklor Narodov Tajmyra. Entsy. Vol. 1. Dudinka.

Labanauskas, Kazimir I. 2002. Rodnoe slovo: ènetskie pesni, skazki, istoričeskie predanija, tradicionnye rasskazy, mify [Mother tongue: Enets songs, tales, legends, traditional stories, myths]. St. Petersburg: Prosveščenie.

Lewis, M. P., Gary F. Simons and Charles D. Fennig (eds.). 2016. Ethnologue: Languages of the world. 19th edition. Dallas: SIL International. (http://www.ethnologue.com/)

Mithun, Marianne. 2014. The data and the examples: comprehensiveness, accuracy, and sensitivity. In Nakayama & Rice (2014, 25–52).

Mosel, Ulrike. 2014. Corpus linguistic and documentary approaches in writing a grammar of a previously undescribed language. In Nakayama & Rice (2014, 135–157).

Nakayama, Toshihide and Keren Rice (eds.). 2014. The art and practice of grammar writing. Honolulu: University of Hawai'i Press.

Shluinsky, Andrey. 2010. "Kontrastivnye" glagol'nye okončanija v lesnom dialekte èneckogo jazyka ["Contrastive" verbal endings in Forest Enets]. In S. I. Burkova (ed.) Materialy 3-j meždunarodnoj konferencii po samodistike [Materials of the 3rd international conference on Samoyedic languages]. Novosibirsk: L'ubava. 279–291.

Siegl, Florian. 2013. Materials on Forest Enets, an indigenous language of Northern Siberia. Helsinki: Société Finno-Ougrienne.

Sorokina, Irina P. 2010. Èneckij jazyk [Enets]. St. Petersburg: Nauka.

Sorokina, Irina P. and Darja S. Bolina. 2001. Slovarj enetsko–russkij i russko–enetskij. St. Petersburg: Prosveščenije.

Sorokina, Irina P. and Darja S. Bolina. 2005. Enetskije teksty. St. Petersburg: Nauka. (http://iling.spb.ru/nord/materia/ency_tit.html, accessed 17 July 2011.)

Sorokina, Irina P. and Darja S. Bolina. 2009. Slovarj enecko-russkij i russko-eneckij [Enets–Russian and Russian–Enets dictionary]. St. Petersburg: Prosvjaščenie.

Susekov, Vasilij A. 1977. Vokalizm èneckogo jazyka (èksperimental'no-fonetičeskoe issledovanie na materiale dialekta baj) [Enets vowels (experimental phonetic study based on the Baj dialect data)]. Doctoral dissertation. Leningrad.

Tereščenko, Natalja M. 1966. Eneckij jazyk [enets]. In V. E. Lytkin and K. E. Majtinskaja (eds.) Jazyki narodov SSSR: Finno-ugorskie i samodijskie yazyki [Languages of the USSR: Fenno-Ugric and Samoyedic languages]. Moscow: Nauka. 438–457.