

Preparing a grammar for Beserman Udmurt

Timofey Arkhangelskiy (timarkh@gmail.com)

Universität Hamburg / Alexander von Humboldt Foundation

Maria Usacheva (mashastroeva@gmail.com)

Institute of Linguistics, Russian Academy of Sciences

Beserman is an endangered unwritten dialect of Udmurt (Uralic > Permic). Its existing descriptions (Teplyashina 1970, Lyukina 2008, 2016) mostly deal with phonology and morphemics. In the talk, we are going to discuss our efforts in preparing a comprehensive Beserman grammar (which is in its initial stages) and present some thoughts on developing grammars for endangered languages in general. This includes primarily collection of data, writing the text and technical infrastructure.

1. Collection of data. Corpora of endangered languages are usually very small and do not contain all the data needed for writing a grammar. One can gather missing data through elicitation, but this method has its limitations. We propose the following workflow:

- formulate hypotheses based on material from corpora;
- conduct experiments to test them;
- use elicitation to gather whatever data remains to be gathered.

We use different experimental methods to get different types of texts: referential communication tasks, postscoring of mute cartoons, telling stories based on a series of pictures, adopting games, etc. The experimental texts are transcribed and sound-aligned in ELAN with the help of the native speakers.

2. Writing the text. When writing any kind of text, it is very important to keep in mind its target audience. Unfortunately, there are several target audiences in our case, who require different data and structure. For example, typologists need plenty of glossed examples and a very detailed text in English; syntacticians need examples (both positive and negative) which illustrate outcomes of syntactic tests; regional linguists from Russia who work with Permic languages need a text in Russian and terminology based on Russian linguistic tradition, etc. As for the Beserman themselves, young speakers need a simple understandable text in Beserman (ideally) or in Russian with plenty of unglossed examples (interpreting glosses is an additional difficulty for non-linguists) and an index of lexemes used in examples which could help them find sentences containing the words they need. Older speakers need a simple text preferably in Beserman with unglossed examples; such a text must be printed, typeset in large characters and include a substantial appendix with photos of the speakers, their traditional activities and a place where they live.

It is a real challenge to satisfy all these target audiences, and we have not found an optimal solution yet. Our preliminary suggestion is to write two texts. The first will be a large detailed grammar in English with plenty of glossed examples (both positive and negative) and international terminology. This text will be addressed to typologists, syntacticians and other linguists who write mostly in English. The second text will be written in Russian with unglossed examples (only positive), with minimum of terminology and with photos. This text will be addressed to Beserman speakers of the older generation.

3. *Technical infrastructure.* We use technical solutions for both collecting/researching the linguistic data and for collaborative preparation of grammar. The former includes the following:

- We developed a morphological analyzer (which has more than 95% coverage) to automatically process transcribed texts. This allows us to save a lot of time when analyzing dozens of hours of experimental recordings, compared to manual analysis in FLEX we performed earlier.
- We put all transcribed texts (both experimental and non-experimental) in a searchable sound-aligned corpus with morphological annotation.
- We keep a separate corpus of usage examples taken from the dictionary we work on, as well as from questionnaires used for researching grammatical phenomena in the field. This corpus includes both positive and negative examples with the possibility of search in both kinds.

To work on the grammar draft, we set up a website with a wiki-engine. This allows us to edit the text collaboratively, to keep the history of changes, and to identify spots in the grammar on which we currently have insufficient data and that have to be researched in the field. This initial draft in wiki-format is designed to comprise all our knowledge in a form most suitable for us rather than the target audiences (e.g. we do not use glosses and store research questions together with the facts that we already know). The texts that we are going to publish will be prepared separately when the work on the wiki-grammar is completed.

References

Lyukina, Nadezhda. 2008. Osobennosti jazyka balezinskix i jukamenskix besermjan (sravnitel'naja xarakteristika) [The features of the language of the Balezino and Yukamenskoye Beserman (comparative characteristics)]. PhD thesis. Izhevsk.

Lyukina, Nadezhda. 2016. Fonetiko-morfologicheskie osobennosti jazyka lekminskix i jundinskix besermjan [Phonetic and morphological features of the language of the Lekma and Yunda Beserman]. Izhevsk.

Teplyashina, Tamara. 1970. Jazyk besermjan [The language of the Beserman]. Moscow

Rethinking the documentation trilogy in endangered language research

Bernat Bardagil-Mas

University of California, Berkeley

Research on languages spoken by minority indigenous communities has long been tied to revitalization initiatives and linguistic documentation. A prototypical approach to such work is the so-called “documentation trilogy”, consisting of a reference grammar, a collection of transcribed and annotated texts, and a dictionary. In this presentation, we put forward the advantages of rethinking the traditional approach into a process with a higher degree of community involvement. Such projects have a long-standing tradition as well, but they have remained a less popular approach than the traditional trilogy.

This presentation discusses one such example in the author’s ongoing linguistic work on Panará (Jê) in Central Brazil to document the language and develop materials for use by the community. A digital dictionary and an alphabet reform are the concrete materialization of the Panará community’s desire and request, which results from (1) the growing distance between the speech of the younger Panará and the village elders, (2) the realization that the older generation, young adults before contact in 1973, will soon disappear, and (3) a feeling that the current orthography is not a satisfactory tool to write the language. Recorded and transcribed texts provide documentation materials on Panará language and culture, while at the same time feed the digital dictionary with lexical entries, contexts of use and recordings. In parallel, workshops and sessions with Panará school teachers and interested members of the community explore the best approach for a faithful and functional spelling system.

The research presented here contains the three elements of the classic documentation trilogy: linguistic research, language documentation and lexicography. However, the way in which they are combined is fine-tuned to the specific needs and interests of the speaker community, resulting in a process with a strong community involvement that still meets the scientific goals of the linguistic research.

Detecting grammatical change in a previously undescribed endangered language

Russell Barlow
 Max Planck Institute for the Science of Human History
 barlow@shh.mpg.de

Key words: structural change, language loss, syntax, morphology,
 Ulwa, Papuan languages, Tok Pisin

While structural change is probably a diachronic reality for all languages (and especially for languages in contact situations), obsolescing languages are likely to experience especially pronounced (and even at times unpredictable) changes to their grammars (Campbell & Muntzel 1989). Although the modern-day grammar-writer may eschew the urge to describe a supposedly pristine form of a dying language—a variety perhaps no longer even spoken—a thorough account of a language in flux would ideally indicate the lexical and grammatical characteristics that are due both to contact and to attrition. Thus, even as we strive to be descriptive (and not prescriptive) of actual contemporary language use, we must attempt to indicate recent grammatical changes—especially when these changes may be attributable to language loss.

This task is especially challenging when the language being described has never been described before. How are we to tease out the structural consequences of language loss, when there exist no earlier records of the language—records that presumably could have revealed the structures of a bygone vital variety of the now decaying language? This descriptive task can prove even more challenging if the language in question is an isolate or if its relatives have not been described either—that is, if no comparative evidence is available.

This presentation focuses on data from Ulwa, a severely endangered language of Papua New Guinea that, prior to 2015, had not been described. Neither of Ulwa's sister languages in the small Ulmapo family has been described either. Ulwa speakers have been shifting over the past few decades to Tok Pisin, the English-based creole that serves as the lingua franca for most of Papua New Guinea. The local language is no longer being transmitted to children, no one younger than 40-years-old is fully proficient in the language, and even the oldest speakers are more comfortable using Tok Pisin than Ulwa. Unsurprisingly, when people do speak Ulwa, they often codeswitch to Tok Pisin. Less immediately obvious, however, are the forms and structures in contemporary Ulwa that have arisen (or have been changed or even lost) due to its obsolescence. Although borrowings from a dominant language are usually easy to identify (especially when they are recent), structural changes can be harder to uncover. Some of these changes may be grammatical calques from the dominant language, but others may be simplifications, omissions, or superfluities caused by imperfect learning or fostered by a reduction in frequency or domains of use.

One aim of this talk is to add to our typological understanding of language loss and grammatical change by providing examples from Ulwa. Many of these changes conform to hypotheses that speakers in endangered language situations tend to rely more on analytic constructions, replacing or reducing the number of their synthetic alternatives (Andersen 1982). At least one suspected change runs counter to claims of the unidirectionality of grammaticalization changes (Hopper &

Traugott 1993), since a historically bound suffix in Ulwa has come to be used as an unbound auxiliary verb. The second aim of this talk, then, is to detail the difficulties that were faced, first in identifying such structural changes in Ulwa, and second in determining which changes have been due to contact and which have been due to language loss.

References

- ANDERSEN, ROGER W. 1982. Determining the linguistic attributes of language attrition. In Lambert, Richard D. and Barbara F. Freed (eds.), *The Loss of Language Skills*, 83–118. Rowley: Newbury House.
- CAMPBELL, LYLE AND MARTHA C. MUNTZEL. 1989. The structural consequences of language death. In Dorian, Nancy (ed.), *Investigating Obsolescence: Studies in Language Contraction and Death*, 181–196. Cambridge: Cambridge University Press.
- HOPPER, PAUL AND ELIZABETH TRAU GOTT. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Challenges constructing a multilectal grammar – the case of the Western Serengeti languages

Rasmus Bernander, Lotta Aunio, Antti Laine & Tim Roth (University of Helsinki)

The purpose of this talk is to report on the challenges (and potential solutions) of producing a digital multilectal grammar over the closely related yet divergent Bantu language varieties of the Western Serengeti subgroup, viz. Ikoma, Nata, Isenye and Ngoreme (Aunio et al. forthcoming).

As pointed out by e.g. Evans (2003), a multilectal grammar comes with numerous advantages. Comparing several closely related - and thus largely similar - linguistic systems can, in those cases where the varieties slightly differ, facilitate the analysis and explanation of synchronic peculiarities as well as the direction of historical processes of change. However, Evans (2003) also rightly notes that there are several challenges connected to compiling a multilectal grammar. One of these problems regards the organization and presentation of the linguistic outcome in those cases where the different language varieties diverge considerably, whether it concerns reflexes of the same form having different functions or the same function being expressed by different forms or even entire subsystems being differently conceptualized.

The aspectual system of the Western Serengeti varieties is an example of a problematic subsystem in this regard, exemplified in (1) with the expressions of the Imperfective and the Progressive respectively in Ikoma and Ngoreme. In this specific instance, the form used for the Imperfective in Ikoma is used for the Progressive in Ngoreme and vice-versa. At the same time, this formal relationship is complex. Thus, for example, the Progressive construction in Ngoreme is both a non-fused but also an inverted reflex of the Imperfective of Ikoma.

(1)		Ikoma	Ngoreme
	Imperfective	<i>N-SM-Vko-B-a</i>	<i>SM-ra-B-a</i>
	Progressive	<i>SM-ra-B-a</i>	<i>N-ko-B-a # SM-V-ní</i>

In addition to such presentational complications, a multilectal grammar struggles with similar problems as those of monolectal grammars, for example, how to make the results reproducible for other researchers (cf. Maxwell 2014) as well as easily retrievable for the actual speakers.

In our making of a multilectal grammar of the Western Serengeti varieties, we are trying to tackle these challenges by the use of a digital platform combining various infrastructural components and state-of-the-art computational tools (developed in cooperation with FIN-CLARIN/Language Bank of Finland <https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/>). We will present this ongoing work, while we also wish to extend this presentation into a discussion on the best practices of creating, containing and distributing a grammar of this kind.

REFERENCES

- Aunio, L. et al. forthcoming. The Mara languages (JE40). In: Bostoen, K & Van de Velde, M. (eds.). *The Bantu languages*. 2nd edition.
- Evans, N. 2003. *Bininj Gun-wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune*. 2 vols. (*Pacific Linguistics*, 541). Canberra: Australian National University.
- Maxwell, M. 2012. Electronic grammars and reproducible research. In Sebastian Nordhoff (ed.), *Electronic grammaticography (Language Documentation & Conservation Special Publication No. 4)*, 207–234. Honolulu: University of Hawai'i Press.

Closing the cycle of language documentation, corpus building and corpus-based description of Komi-Zyrian

Our paper presents a project aiming at both an annotated corpus and a corpus-based syntactic description of Komi-Zyrian, an endangered Uralic language. It is relatively vital and actively used in many spheres of modern life, although definitely endangered due to heavy language shift to Russian.

Many projects working with the documentation and description of endangered languages start recording and transcribing texts, manually tagging them for morphosyntactic categories, before analysing the resulting corpus and writing a typologically informed descriptive grammar. Several such grammars have recently been announced under the label “corpus-based” (e.g. Wilbur 2015), although these grammars only use examples from the corpus for argumentation and illustration, rather than actually testing the hypothesis about the grammar of the relevant language based on the corpus.

One way to carry out corpus-based testing of grammar predictions (and linguistic hypotheses in general) is the implementation of a grammatical description in a computer-processable format, i.e. as a corpus parser. Since such a description must be highly formalised, it differs fundamentally from typologically informed descriptive grammars traditionally written in language documentation projects.

We present a workflow that combines language documentation with incremental corpus-based language description, including automatised corpus annotation. Our work relies on the open Natural Language Processing (NLP) infrastructure available for Northern Eurasian endangered languages at Giellatekno (Trosterud 2006). Whereas Giellatekno has thus far been working only with written language texts, our workflow begins using speech recordings with aligned transcriptions in ELAN. Finite-State morphology is used for rule-based modeling of stems and segmental affixes, as well as complex morphophonological rules. Additional rules following Constraint Grammar (Karlsson et al. 1995) are implemented for syntactic disambiguation and tagging dependency relations. The source code and documentation is being developed using a versioning system and is available under open license.

Our approach might not be applicable to the majority of the world’s endangered languages, for which sufficient linguistic description is not available. However, as regards basic description and standardisation, many endangered languages of developed countries are comparable to Komi-Zyrian. In Russia, for instance, this is true for various Uralic, Turkic, and other languages. With regard to northern Eurasian languages specifically, we want to challenge the current mainstream in the field, which relies on time-consuming manual (and occasionally semi-manual) morphosyntactic annotation during corpus-building. We question the usefulness of investing resources in manual interlinearisation and glossing of already known grammatical features. Instead, we want to invest effort in creating more automated methods, whilst linking documentary linguistics more closely to established NLP methods.

Our approach is novel in documentary linguistics, which has only rarely considered applying NLP to build annotated corpora more efficiently. Although fundamentally different from established grammaticography in Documentary Linguistics (cf. Mosel 2006), we argue for the high potential of this approach in the documentation of endangered languages as well as in creating language technology for language production and teaching. Last but not least, such a grammar can also be the basis of a less formal descriptive grammar that is ideally based on the same annotated corpus.

References

- Karlssohn, Fred; Arto Voutilainen, Juha Heikkilä & Arto Anttila (eds). 1995. *Constraint Grammar: A language-independent system for parsing unrestricted text*. Berlin: Mouton de Gruyter.
- Mosel, Ulrike. 2006. Grammaticography: The art and craft of writing grammars. In: *Catching language: The standing challenge of grammar writing*, ed. by Felix Ameka, Alan Dench & Nicholas Evans. Berlin: Mouton de Gruyter, 41–68.
- Trosterud, Trond. 2006. Grammatically based language technology for minority languages. In: *Lesser-known languages of South Asia: Status and policies, case studies and applications of information technology*, ed. by Anju Saxena & Lars Borin. Berlin: Mouton de Gruyter, 293–316.
- Wilbur, Joshua. 2015. *A Grammar of Pite Saami*. Berlin: Language Science Press.

‘Indefinite’ or ‘antipassive’? Matching diachronic and synchronic approaches in the study of Ainu

Anna Bugaeva, Tokyo University of Science/NINJAL

bugaeva@rs.tus.ac.jp

Ainu (isolate), a nearly extinct language, originally spoken in Northern Japan (Hokkaido, Northern Honshu) and the Far East of Russia (Southern Sakhalin, Kuril Islands, and presumably Southern Kamchatka), presents many challenges for describing it in its own terms and, at the same time, making it comparable to other languages. In fact, Ainu shares few areal features with Northeast Asian languages, i.e. Japanese, Korean, and the so-called “Altaic” languages. It is a polysynthetic, incorporating and predominantly head-marking language with mixed alignment (nom-acc, neutral and tripartite) and quite complex verbal morphology comprising a wealth of voices, verbal number and other features which are rather reminiscent of North American languages.

This paper focuses on the Ainu-style antipassive *i-* (Bugaeva 2004), which has traditionally been regarded in Ainu grammars as the indefinite object marker ‘(indefinite) person/thing’ (Tamura 1988: 67) and therefore is missing in *WALS* (2005). This paper clearly shows that there are two synchronically distinct *i-* markers in Ainu, viz. the derivational antipassive *i-* and inflectional ‘fourth’ person object *i-* with the functions of first person inclusive (1PL.INCL), second person honorific (2HON), and logophoric (person of the protagonist). The so-called indefinite marker *i-* ‘person/thing’ can be regarded as an antipassive marker *per se* based on its syntactic (eliminating a patient/theme/recipient argument), semantic (denoting an unspecified generic participant or lexicalizing it to a single or subset of objects) and discourse (patient-defocusing) properties.

Unlike other affixes of the ‘fourth’/indefinite series, i.e. the personal affixes *a-* for A and *-an* for S, the object prefix *i-* in its proper indefinite/unspecified function is not used as an inflectional object marker. Instead it blocks an overt expression of the object and decreases verbal valency (1b). Thus antipassives from bivalent transitives result in intransitives, which is additionally evidenced by the change of the transitive subject indexing *ci-* in (1a) to the intransitive subject indexing *-as* (1b). Generally, the antipassive construction (1b) is preferred when there is a focus on A and the action itself rather than on the object participant (P).

(1) a. *menoko yukar ci-nu kor oka-as*
 woman epics 1PL.EXCL.A-hear when exist.PL-1PL.EXCL.S

‘We are listening to female epics (=songs of gods).’ (Tamura 1987: 79)

b. *orwa i-nu-as hike*
 then ANTIP-hear-1PL.EXCL.S according

‘(We saw the flocks of birds.) Then from (what) we heard...’ (Kubodera 1977: 424)

Contrary to the accepted view, I adduce the ‘antipassive to 1PL.INCL.O’ scenario based on extensive crosslinguistic and Ainu-internal evidence. The antipassive *i-*, in its turn, originated in the incorporation of a generic noun **i* ‘thing/place/time’, which is not unusual in languages without overt expression of the demoted O participant in the antipassive. The extended use of the antipassive *i-* is attested on obligatorily possessed nouns to enable their use without possessive affixes, which is similar to ‘absolutive’ derivational markers attested in Uto-Aztecan, while in the Athabaskan languages the same function is performed by the unspecified possessor affix, which is part of inflectional paradigm.

Synchronic misinterpretations and inaccuracies in descriptions of certain categories in particular languages are often rooted in the diachrony of respective markers, but this does not mean that the diachrony as such should be neglected in grammar writing. Both synchronic and diachronic approaches are equally important for linguistic description and typology, it is just important to clearly separate them.

References

- Bugaeva, Anna. 2004. *A grammar and folklore texts of the Chitose dialect of Ainu (Idiolect of Ito Oda)*. ELPR Publication Series A-045. Suita: Osaka Gakuin University.
- Kubodera, Itsuhiko. 1977. *Ainu bungaku* [Ainu literature]. Tokyo: Iwanami Shoten.
- Polinsky, Maria. 2005. Antipassive Constructions. In Martin Haspelmath, Matthew S. Dryer and David Gil (eds.), *The World Atlas of Language Structures*, 438–439. New York: Oxford University Press.
- Tamura, Suzuko. 1987. *Ainugo onseishiryō 4* [Ainu Audio Materials 4]. Tokyo: Waseda daigaku gogaku kyōiku kenkyūjo.

How historical syntax helps the description of uncommon grammatical features
Natalia Cáceres Arandía (University of Oregon)

In this paper I give an overview of the challenge in describing four grammatical structures in languages of the Cariban family. I argue that by tracing the source of the syntactic constructions that gave rise to these structures, it is possible to integrate the “anomalous” patterns in the relevant typological discussions.

The first “quirk” of the grammar of Cariban languages came into the spotlight of typology when Derbyshire (1979) described Hixkaryana as having basic OVS order. Other features of the grammar that differ from what has been described in the typology include:

- (i) an imperfective ergative split (Sapién 2017; Cáceres Arandía 2015)
- (ii) an intransitive split that has no semantic basis (Meira 2000)
- (iii) several clauses functionally competing in the domain of passive voice but not exclusively dedicated to this function (Gildea 2014; Cáceres Arandía 2017)
- (iv) a separate part of speech for property concepts that cannot be used as an adnominal modifier but as a predicate modifier or a secondary predicate (Meira and Gildea 2009; Cáceres Arandía 2015)

Taken synchronically, each of these features would have to be considered marginal cases in the realm of possible grammatical systems. However, an analysis in terms of constructions with layers of history shaping their synchronic form and function, allows each feature to be understood against typologically well-documented paths of evolution.

Thus, the imperfective ergative split is just another case of reanalysis of an action nominalization (*the destruction of the city*) with the addition of a *by*-clause for expressing the agent (*by the enemy*) into a main clause (*the enemy destroys/is destroying/will destroy the city*); the intransitive split is accidentally introduced as an affixal valency reducing strategy (*to kill O-> kill self/be killed*) which yields intransitive meanings that lexicalize progressively replacing part of the original basic intransitive verb inventory (*die = kill self/be killed*); the passives originate from well-known sources (reflexives, resultative participles, patient nominalizations), and, if several of the sources are available in one language, speakers have various options; finally, the unusual distribution of parts of speech is the result of the development of a class of words specialized in the function of secondary predication.

As Rankin (2006) points out, our field insisted on keeping synchronic description separate from comparative and historical analyses, but there are cases in the grammar of Siouan languages that can only make sense in a comparative and historical light. Remembering the latter and taking the characteristics of whole constructions into account instead of considering morphemes outside of their morphosyntactic contexts are crucial steps when describing the grammar of each language of the Cariban family.

References

Cáceres Arandía, Natalia. 2015. The Ye'kwana ergative imperfective: a case of reanalysis. *International Workshop on Diachronic Morphosyntax in South American Languages*, Lyon, Laboratoire Dynamique du Langage. May 28-30

- Cáceres Arandia, Natalia. 2015. Constructing Cariban property concepts through adverbs, *ALT 2015 The 11th meeting of the Association for Linguistic Typology*, University of New Mexico, August 1-3
- Cáceres Arandia, Natalia. 2017. Finding passives in Ye'kwana (Cariban): too many might still not be enough, *50th Annual Meeting of the Societas Linguistica Europaea*, Zurich, September 10-13, 2017
- Derbyshire, Desmond C. 1979. *Hixkaryana*. Amsterdam: North-Holland.
- Gildea, Spike. 2014. Diachronic Typology of Passive in the Cariban Family. *Workshop: Voice systems in diachrony: a comparative perspective*. University of Pavia, Italy, 11 September
- Meira, Sérgio. 2000. The Accidental Intransitive Split in the Cariban Family. In Spike Gildea (ed.) *Reconstructing Grammar: Comparative Linguistics and Grammaticalization*, 201–230. *Typological Studies in Language* 43. Amsterdam, Netherlands: Benjamins.
- Meira, Sérgio, and Spike Gildea. 2009. Property concepts in the Cariban family: adjectives, adverbs, and/or nouns. In *The Linguistics of Endangered Languages- Contributions to Morphology and Morphosyntax*, ed by. Leo Wetzels, 95–133.
- Sapién, Racquel-María. 2017. An Innovative Main Clause Construction with Ergative Patterns in Kari'nja (Cariban, Suriname). *Internationa Journal of American Linguistics* 83(2): 207–231

Verb agreement in Darai and Majhi (Nepal) in Areal-Typological Context

Dubi Nanda Dhakal (Central Department of Linguistics, Tribhuvan University, Kathmandu, Nepal)

This paper explores the verb agreement in two minor Indo-Aryan languages spoken by small communities in the eastern Nepal. Darai (iso dry) and Majhi (iso mjz) share a number of typologically interesting features, such as head marking in genitive construction. The head of the noun phrase (NP) modified by the genitive-marked noun, and the nouns that appear as heads also take the pronominal possessive suffixes. Both of these languages have the ergative subjects based on nominal hierarchy.

The verb agreement with the gender, person, number, and honorificity is common across Indo-Aryan languages of Nepal. Interestingly, both the subject, and object are coded in the transitive and ditransitive sentences in these two languages. It is common in other languages of the region, such as Maithili and Rajbanshi (Wilde 2007) as well. Let's consider example (1) from Darai in which both the subject, and object are coded in the verb. By contrast, if the object is the third person singular, the finite verb will be *de-la-mi-k* 'give-PST-1SG.SUB-3SG.OBJ'.

- (1) *məi terake ʈaŋgi delamir*
 məi toi-ke ʈaŋgi de-la-m-r
 I you-DAT axe give-PST-1SG.SUB-2SG.OBJ
 'I gave an axe to you.' [Darai]

Despite the fact that both the subject and object information is coded in the transitive and ditransitive clauses in Majhi, Majhi makes use of the portmanteau suffix to code both the subject and object as in (2).

- (2) *muĩ tshoḍarilai dzal dinin*
 muĩ tshoḍari-lai dzal di-n-in
 I son-DAT net give-PST-1SG.SUB.3SG.OBJ
 'I gave a net to my son.' [Majhi]

If the object is the second person singular, the verb form will end as *di-la-i* 'give-PST-1SG.SUB.2SG.OBJ'.

A less but typologically interesting features in these languages is the verb agreement with dative-marked subjects in stark contrast with some major Indo-Aryan languages do not have this feature. The agreement marker *-e* is attached to (3) in Majhi, agreeing with the first person singular. If the subject is the second person singular, the verb will inflect as *-jas* '2SG.DAT'. This can be extended to the non-past tense as well.

- (3) *milai bhok lagle*
 muĩ-lai bhok lag-l-e
 I-DAT hunger feel-PST-1SG.DAT
 'I was hungry.' [Majhi]

Similar to example (3) in Majhi, Darai also shares similar pattern. The verb agreement is also triggered by dative subject taking distinct suffixes in Darai.

The areal features associated with the verb agreement discussed here can be instrumental in analyzing the verb agreement with the minor Indo-Aryan languages spoken in the same geographical region. An Indo-Aryan language, such as Danuwar (Kuegler and Kuegler 1974) seems have the same kind of features, but its grammar is yet to appear. Taking consideration of the above facts, it can be said that the areal typology becomes an instrumental in describing the grammatical features of the lesser-described languages in Nepal.

References

- Kuegler, Klaus P. and Doris Kuegler. 1974. *Danuwar Rai Phonemic Summary*. Kathmandu: Summer Institute of Linguistics.
 Wilde, Christopher P. 2008. *A Sketch of the Phonology and Grammar of Rajbanshi*. Helsinki: University of Helsinki.

Event structure in Tomo Kan Dogon: adopting the formal approach in a descriptive grammar
Vadim Dyachkov
Institute of Linguistics, Russian Academy of Science

The paper deals with formal aspects of describing the aktionsart and aspect in Tomo Kan language (< Dogon < Niger-Congo), especially with the problem of how Wendler's classes of predicates can be distinguished in the language. It can be shown that standard tests do not allow to draw the line between activities and accomplishments in Tomo Kan since all the verbs are compatible both with *in-* and *for-* adverbials or only with *in-* adverbials (3):

- (1) ń sá ńá: mínitì pyèlè: (wá) ńyèⁿ
 1SG REFL food.H minute 10 LOC eat.PFV
 'I ate my food for (in) 10 minutes'.
- (2) ń dàʔàǰì ló: (wá) wá:ńá bír-è: (ń dùmà=ⁿ)
 1SG day 2 LOC work.H do-CH 1SG finish=PFV.NEG
 Lit. 'I worked for (in) two days (but I haven't finished yet)'.
- (3) ǰì: mìnìtì pyèlè: *(wá) bèmì
 water minute 10 LOC boil.PFV
 'The water boiled in (*for) 10 minutes'.

I suggest that the problem can be easily solved if the semantics of the Tomo Kan Perfective is correctly defined. I will show that the (in)compatibility of a verb with *in-* and *for-* adverbials can be predicted based on the formal definition of Perfective semantics and extralinguistic knowledge of the situation described by a verb.

In order to account for the observed facts I will use Ramchand's event structure framework (cf. Ramchand 2008) according to which a verb can be decomposed into one to three subevents. However, I will try to show that Ramchand's theory cannot be applied in a straightforward way to the Tomo Kan data and that additional components of description must be added. At the same time, adopting Ramchand's approach will also facilitate description of other components of the grammar such as causative derivation. It turns out that morphemes which have cognates in other Dogon languages (see Plungian 1993) derive causatives from verbs (4) and inchoatives from adjectives (5):

- (4) ń ńkó sùbá àbà-rí
 DEM DEM 1SG.TOPIC be.spread-CAUS.PFV
 'I spread the mat'.
- (5) kò ʔwà: nòmbò-rí
 DEM well high-CAUS.PFV
 'The well has become deeper'.

I argue that causative markers occupy the procP syntactic position, and their function is to introduce a process sub-event. This sub-event denotes the transition to a resulting state but not the agent's efforts. Causative interpretation arises directly from the semantics of the stem and can be described as procP-to-initP movement (consider also McIntyre 2010's BECOME=CAUSE hypothesis).

The grammar of Tomo Kan (in preparation) is intended to be descriptive and cannot include any references to formal semantics and/or generative linguistics since it is addressed to specialists which may be not interested in formal aspects of the phenomena in consideration. I will focus primarily on how the event structure is created in Tomo Kan language and how the aktionsart and derivational morphemes are described in the grammar. Firstly, I will describe the machinery of detecting aktionsart classes in Tomo Kan. Secondly, I will describe the semantics of causative markers as well as other similar affixes (namely reversives) and show that their ability to serve as labile predicates crucially depends on the semantics of a verbal/adjectival stem to which they are attached. Thirdly, I will show how all the observations made are reflected in the text of the grammar.

The paper is based on the author's own fieldwork (including several trips in 2011 – 2018), primarily on the author's own questionnaires and corpus of Tomo Kan texts collected during the fieldwork.

The research has been supported by Russian Science Foundation, grant No. 17-78-20071.

References

McIntyre 2012 – McIntyre, Andrew. 2012. The become = cause hypothesis and the polysemy of get. *Linguistics*, 50(6), 1251 – 1287.

Plungian 1993 – Plungian, Vladimir. 1993. Three causatives in Dogon and the overlapping of causative and passive markers. In Comrie, B. et al (ed.), *Causatives and transitivity*, 391 - 396. Amsterdam: Benjamins.

Ramchand 2008 – Ramchand, G. *Verb Meaning and the Lexicon*. Cambridge University Press, 2008.

**The Mochica language description according to a colonial grammar
(17th century)**

Rita Eloranta (Hanken School of Economics/University of Leiden)

Most of the languages that existed prior to the Spanish conquest vanished, leaving little evidence behind. Mochica is an extinct linguistic isolate that was spoken until the mid-to late-nineteenth century in the northern coastal area of Peru. It is typologically distinct from other Andean languages. Mochica is predominantly a synthetic, suffixing language. Fortunately, in the case of Mochica, there are colonial and post-colonial descriptions which help us interpret some of its forms. The Mochica language has been preserved whilst it was still spoken in *Arte de la lengua yunga* by Fernando de la Carrera (1644). Our knowledge of the grammar and phonology of Mochica is limited since Carrera's (1644) is the only existing grammar that has survived since colonial times.

Carrera's is a typical colonial missionary grammar that offers problems for linguists that want to understand and describe this language. In colonial grammars languages were described following the Greco-Latin grammatical tradition. In this presentation I aim to first address some of the difficulties colonial grammars of Spanish tradition offer, in general. Secondly, I will present the difficulties I had to face when describing this language. I will discuss about (a) case (morphological and adpositional) in Mochica, (b) its peculiar numeral classifier system (comparable with the ones present in Polynesian languages!) and (c) Mochica nominal possession.

Carrera, Fernando (1644) *Arte de la lengua yunga*.

Writing a grammar for an Amazonian languages

Sidi Facundes, Universidade Federal do Pará
 Marília Fernanda P. de Freitas, Universidade Federal do Pará
 Bruna Fernanda S. Lima-Padovani, Universidade Federal do Pará
 Brazil

Our talk will examine the challenges that Amazonian languages posit to grammar writing and linguistic typology, in light of the linguistic diversity of the region. The Amazon region is well known for a rich diversity in flora, fauna, as well as in the traditional culture of its indigenous peoples. Less known is its extremely rich linguistic diversity. While Europe's languages can be traced back to a single ancestral language, Proto-Indo-European, the linguistic diversity in the Amazon region can only be explained if several ancestral languages (such as Proto-Arawak, Proto-Macro-Jê, Proto-Tupi, among others) are posited. Such languages diversity means typological diversity and can be well illustrated by Apurinã, spoken in Northwestern Amazon, in Brazil, a language of the Arawak linguistic family, the family to which also belonged Taino, the language Christopher Columbus had the first contact with in the New World. Such challenges can be illustrated by what Facundes 1994 and Facundes 2000 described as a number of bound "floating" morphemes in Apurinã, which are phonologically bound formatives with some word-like grammatical properties. As bound morphemes that have clear affix positions in the word forms, they resemble affixes, but their semantic or functional scope as well as the possibility of attaching to different word classes and in different places in the clause make them resemble full word forms. By way of illustration, a subject/possessor person marking for first person *ny=* is almost equivalent to its independent corresponding form *nuta*, although the first can function only as subject and possessor, such as in *ny=myteka* 'I ran' and *ny=kywy* 'my head', whereas the second can also function as object: *nuta myteka* 'I ran', *nuta kywy* 'my head', and *0-akirita-nu nuta* [3Subj=call=1Obj.Sg 1Sg 'He called me'. Different from cross-reference marking in other languages, these person markers can occur alone in the sentence without their co-referential noun phrase, but they normally cannot co-occur with preverbal co-referential noun phrases, when used in verbs. That is, such person markers behave sometimes like cross-referencing markers, other times like

bound noun phrases. In relation to tense or aspect markers, these are special insofar as they can occur attached to nouns, pronouns or verbs, regardless of their semantic scope. So, *nuta=ku apuka* and *nuta apuka=ku* both mean 'I'll arrive'. Finally, some oblique case-marking functions such as associative are marked by bound morphemes that always attach to the right edge of a noun phrase. So, we have *nuta=kata* 'with me' and *nuta ytanyru ynyru=kata* (1Sg wife mother=ASSOC) 'with my wife's mother'. In our talk, we will discuss the issues that such grammatical features present to grammar writing and typology, and how they can be addressed in the Apurinã language grammar.

References Cited

- Facundes, Sidney da Silva. 1994. Noun Categorization in Apurinã (Maipuran). Unpublished MA. Thesis presented to the Linguistic Department of the University of Oregon, Eugene, Oregon.
- _____. 2000. The Apurinã Language of Brazil (Arawak). PhD Dissertation: SUNY-Buffalo, Buffalo, NY.

The distribution of articles in three Otomanguean languages

H. Antonio García Zúñiga, Instituto Nacional de Antropología e Historia
Armando Mora-Bustos, Universidad Autónoma Metropolitana
Alejandra Itzel Ortiz Villegas, Instituto de Educación Media Superior

The aim of this paper is analyze the distribution of the article in certain syntactic positions and the possibilities of combination with other referential elements, such as demonstratives in three languages of the Otomanguean family (Mexico): Mazatec, Mazahua and Amuzgo. These three languages have been selected because they fulfill the first division of the wide Otomanguean family: the Mazahua belongs to the western division, while the Amuzgo and Mazatec belong to the eastern division; these otomanguean languages present less dialectal diversity.

In Mazatec, the entity referred to in the nominal phrase is identifiable even when that phrase is not accompanied by an article (1a). If a noun phrase is modified by this mark, the entity acquires the property of specificity (1b). The interesting thing in this language is that the article can co-occur with a demonstrative, (1c), which expresses the deictic anchor and indicates the perceptual space between the speaker and the listener, as well as the distance between them.

- 1 a. na²hpo² tse¹ ʃõ¹ niq³ hʃa³ k^hi³
clothes PREP woman INTEN old COP.SG
'the lady's clothes are very old' (Txt.13)
- b. he chjota xrin sinda choa+nkjain
ART people man make.3 container+face.3
'The gentlemen made their masks' (Txt.16)
- c. he³ ʃõ¹ ʃokhi ti³mę:¹³
ART woman DEM get.sick.3
'This woman is sick' (Txt.05)

In Mazahua, there are two sets of articles. In the first one, (2a), there are the articles (nu =, jo =) that have a recognition value of the reference of the nominal and, in the second, (2b), the forms appear (k'i =, k'o =), which express different degrees of definiteness. Only the linked morphemes of the second group have their respective free forms (k'inu, k'oyo), (2c). Now, the language has two forms (= nu, = jo) that code for spatial proximity deixis, and two others that code for distance spatial deixis (= k'i, = k'o), what is striking is that, in formal terms, these marks are the same as those that are related to the definition, only that they occupy a different position in the sentence (2d).

- 2 a. mù=rà=tõnho nù=ts'í-ns'í k'ì=Ø=ʃõʃi a=ʃes'e
when=3FUT=sing ART.REC=DIM-bird SUB=3PRS=stand.up LOC=up
'When the bird, that normally steps up, would sing' (Txt.al.11)
- b. mbò Ø-ndĩnrĩk'ì=p^hunto mí-²jôo=ni
then 3.PST-say ART.DEF=dead 1POS-father=RPT
'when my deceased father said' (Txt.pt.11)

- c. mbò Ø=ndèβe k'ìnu físk^wama=k'ì mì=pès'ì ángeze
 then 3PST=drop ART.DEF papper=SUB 3COP=keep PRON3
 'then the piece of paper he kept fall down' (Txt.sp.12)
- d. fò=ndè=Ø=ts'ót'ì=k'ì; Ø=ndà-ts'ót'ì jò=pàle=k'o
 also=half=3PST=lock=DEM.PROX2 3PST=all-lock ARTPL=grandfather=DEMPL.PROX2
 'Also someone locked him, someone locked those grandparents' (Txt.pt.11)

In Amuzgo, the marking of definiteness, apparently, is simple, a suffix is used for this propose, which is formed by a single vowel whose quality and timber are the same as the preceding syllable, (3a). This morpheme, which is equivalent to a definite article, delimits the noun phrase, (3b). Some nominals, such as those that play a locative role, (3c), or some other adjuncts do not require this morpheme to establish the notion of definiteness.

- 3 a. ø-tyen-an³⁴ besana n-dja-an³⁴ ø-tso³ tyosu=u'
 PST-go-3.PL.RESP nothing.else PST-fight-3.PL.RESP PST-say rabbit=ART.DEF.SG
 'they were to fight, answered the rabbit' (Txt.Fox)
- b. u^{'34}=n³-jo³⁴ ø-yon³⁴ sku^{'12} tyondye=e'
 2PRON=PRS-come.1SG PRS-bring.2SG wife fox=ART.DEF.SG
 'you stole the fox's wife' (Txt.Fox)
- c. kwi=t'i¹²=tyondye ø-ts'a sian jndë¹² tyu'a
 ART.INDEF=partner=fox PRS-do work field early
 'a fox works early in the field' (Txt.Fox)

The importance of this work lies in the fact that, in Mexican languages, has not been yet extensively studied the structure of the nominal phrases where articles are used and these co-occurred with demonstratives. Likewise, at a nominal phrase level a lot of restrictions appeared for the articles applicability which are blocked or activated with locatives, possession marks, focus or topic marks, deictic functions, grades of high or low referentiality. That is why it is necessary to review in detail a hierarchy of features within a scale of definiteness generally expressed through the articles.

REFERENCES

- BECKER, Laura. 2018. *Articles in the world's languages*. Ph.D. diss., Universität Leipzig.
- DRYER, Matthew S. 2013. Definite articles. En Matthew S. Dryer y Martin Haspelmath (editores). *The World Atlas of Language Structures Online*. Leipzig: Mx Planck Institute for Evolutionry Anthropology. <http://wals.info/chapter/37>.
- GIVÓN, Talmy. 2011. *Ute reference grammar*. Amsterdam: Benjamins.
- HASPELMATH, Martin. 1999. Explaining Article-Possessor Complementarity: Economic Motivation in Noun Phrase Syntax. *Language* 75(2). 227–43.
- LYONS, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

**Grammars as overlapping repertoires: Accommodating individual variation in language
description using a community of practice framework**

Sonja Gipper, University of Cologne

Katja Hannß, University of Cologne

There is an ever growing amount of evidence for the empirical significance of the idiolect for linguistic theory (e.g. Johnstone 1996, Barlow 2013). Most fieldworkers will have made the experience that grammaticality judgments vary from speaker to speaker (e.g. Fillmore 1979, Ross 1979, Henry 2005). Le Page & Tabouret-Keller (1985: 193) use the term ‘overlapping repertoires’ to capture the observation that in fact, no two speakers share exactly the same linguistic resources. Accounting for linguistic variation is hence an important enterprise for descriptive linguists (e.g. Nagy 2009). In this paper, we wish to contribute to this endeavour by exploring the usefulness the community of practice framework (Lave & Wenger 1991, Wenger 1998) for language description. While this framework has found ample application in different strands of linguistics such as variationist sociolinguistics (e.g. Bucholtz 1999, Eckert 2000), multilingualism research (e.g. Hall, Cheng & Carlson 2006), and second language learning (e.g. Fraga-Cañadas 2011), it does not form part of the standard toolkit in descriptive linguistics. The groups that speak underdescribed languages are usually portrayed as ‘speech communities’ where people “use the same set of speech-signals” (Bloomfield 1933: 29). However, we argue that it is much more fruitful to analyse these groups as communities of practice. Communities of practice are concrete groups of people who engage in a joint enterprise (Wenger 1998: 73). This framework allows for a flexible view of community: Rather than predefining a group on the basis of a shared language, it focuses on concrete practices without previous assumptions about who will be considered a member. We furthermore suggest contrasting the community of practice with a more abstract notion of speech communities as imagined identities (cf. Anderson 1991 [1983]) that can be evoked through linguistic practice. On our account,

communities of practice are the locus of the negotiation of identity and the emergence of overlapping linguistic repertoires. This twofold approach has a variety of advantages: First, it is useful for accommodating individual and other kinds of variation in language description because the approach puts the individual speaker in focus and sees the overlapping repertoires as emerging from use. Second, it allows us to investigate how speakers construe the linguistic resources they use as belonging to imagined and abstract speech communities, thereby paying attention to local conceptions of language. Third, we are able to account for both the regular component of language, the overlapping repertoires of individual speakers usually called ‘grammars’, and the ultimately creative component that assigns new meanings to familiar linguistic forms every time they are used. We exemplify the usefulness of our approach with data from contemporary documentation corpora of two Bolivian indigenous communities who self-identify as speaking Chipaya and Yurakaré (both isolates).

References

- Anderson, Benedict. 1991 [1983]. *Imagined communities: Reflections on the origin and spread of nationalism*. London/New York: Verso.
- Barlow, Michael. 2013. Individual differences and usage-based grammar. *International Journal of Corpus Linguistics* 18(4): 443-478.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt.
- Bucholtz, Mary. 1999. “Why be normal?”: Language and identity practices in a community of nerd girls. *Language in Society* 28: 203-223.
- Eckert, Penelope. 2000. *Linguistic variation as social practice*. Malden: Blackwell.
- Fillmore, Charles. 1979. On fluency. In Charles J. Fillmore, Daniel Kempler & William S-Y. Wang (eds.), *Individual difference in language ability and language behaviour*. New York/San Francisco/London: Academic Press, pp. 85-101.

- Fraga-Cañadas, Cynthia P. 2011. Building communities of practice for foreign language teachers. *The Modern Language Journal* 95(2): 296-300.
- Hall, Joan Kelly, An Cheng & Matthew T. Carlson. 2006. Reconceptualizing multicompetence as a theory of language knowledge. *Applied Linguistics* 27(2): 220-240.
- Henry, Alison. 2005. Idiolectal variation and syntactic theory. In Karen P. Corrigan & Leonie Elise Alexandra Cornips, *Syntax and variation: Reconciling the biological and the social*. Amsterdam/Philadelphia: John Benjamins, pp. 109-122.
- Johnstone, Barbara. 1996. *The linguistic individual: Self-expression in language and linguistics*. Oxford: Oxford University Press.
- Lave, Jean & Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Le Page, R.B. & Andrée Tabouret-Keller. 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.
- Ross, John Robert. 1979. Where's English? In Charles J. Fillmore, Daniel Kempler & William S-Y. Wang (eds.), *Individual difference in language ability and language behaviour*. New York/San Francisco/London: Academic Press, pp. 127-163.
- Wenger, Etienne. 1998. *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.

What ‘contact typologists’ want from grammatical descriptions

Eitan Grossman and Steven Moran
Hebrew University of Jerusalem and University of Zurich

Language contact is often dealt with anecdotally, in terms of a single contact situation between particular languages of interest. However, it can also be studied with the tools of typology (see, e.g., Seifart 2013, 2015; Haspelmath & Tadmor 2009). For typologists interested in the typology of language contact-induced change (‘contact typologists’), descriptive grammars are a crucial source of information. However, most (if not all) descriptive grammars have lacunas in information about contact situations and contact-induced lexical and structural change. Fortunately, most of these lacunas are fixable, if descriptive linguists are interested in doing so.

In this talk, we give a list of practical suggestions for grammar writers, based on ongoing research on contact-typological projects dealing with phonology (borrowed segments), on the basis of Grossman et al. 2018, and morphosyntax (borrowed case markers and adpositions). In the domain of the typology of borrowed segments, we discuss the difficulties that grammar writers have in their synchronic analysis due to their perceived marginality, whether in terms of frequency or distribution. We suggest standards for treating such segments.

In the domain of case marker and adposition borrowing, we discuss different types of morphosyntactic integration, and suggest general guidelines for presenting borrowed grammatical items in grammars. Lastly, we discuss general desiderata for characterizing multilingual situations in grammars, and for the presentation of examples and texts.

References

- Grossman, Eitan, Dmitry Nikolaev & Steven Moran (eds.). 2018. *Phonological segment borrowing: a worldwide survey*. Jerusalem and Zurich.
- Haspelmath, Martin & Uri Tadmor (eds.). 2009. *The World Loanword Database (WOLD)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wold.cld.org/>.
- Seifart, Frank (ed.). 2013. *AfBo: A world-wide survey of affix borrowing*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://afbo.info/>.

Seifart, Frank. 2015. Direct and indirect affix borrowing. *Language* 91(3). 511–532.

Riho Grünthal (University of Helsinki)

To be or not to be in Veps: reproducing the lack of copula in stative relational clauses

Comparing various branches of the Uralic languages, one of the striking differences is that most languages belonging to the two western-most branches Saamic and Finnic display a copula, whereas more eastern languages have more specific rules for the use or lack of a verb denoting ‘be’ in various syntactic environments. While other Finnic languages regularly manifest a copula in all stative relational clause types (cf. Payne 1997: 114–128), colloquial Veps often diverges from this characteristic and lacks a copula in attributive and equative sentences (1). Literary Veps displays the more wide-spread Finnic pattern (2).

(1) *soba-d redukha-d*
cloth-PL dirty-PL
‘The clothes are dirty.’

(2) *nece om muzeja-n pämez'*
this be.3SG museum-GEN leader
‘This is the leader of the museum.’

The described variation in Veps stative relational clauses affects both affirmative and negative clauses. Syntactically, the more limited use of a copula for tense and mood marking as well as existential clause types resembles, on the one hand, the use of ‘be’ in more eastern Uralic languages and, on the other hand, Russian, the most significant contact language of Veps. Despite the lack of a copula in the Uralic languages, the pattern characteristic of Veps most likely results from language contact with Russian.

From the perspective of grammatical description the varying use and lack of copula in Veps is not limited to stative relational clauses. In dialectal data, a secondary negative existential *iile* ‘there is not’ arising from NEG + ‘be’ is widely used. The most significant effect of the functional reanalysis of the copula, however, affects the description of assumed compound past tenses that Veps literary language displays following the model of other Finnic languages. In vernacular data, participle-based past tense forms characteristic of compound past tense forms such as Finnish (3) and literary Veps (4) connecting a copula with a participle, the copula is typically lacking in most Veps dialects (5).

(3) *hän on tul-lut tänne*
(s)he be.3SG come-PST.PTCP here
‘(S)he has come here.’

(4) *Kazahstana-späi om tul-nu tänna*
Kazakstan-ELA be.3SG come-PST.PTCP here
‘(S)he has come here from Kazakstan.’

(5) *mužik lähtke-n päle kole-nd*
man well-GEN upon die-PST.PTCP
‘The man has died on the well.’

The paper has a bifold focus. Firstly, it will briefly examine the use of copula in Veps stative relational clauses. Secondly, assesses the characteristics of copula clauses in terms of the description of past tens categories in Veps.

Reference

Payne, Thomas 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge: CUP.

Low frequency items in descriptive grammars: A case study of epistemic discourse markers

Karolina Grzech
Stockholm University

This talk is concerned with the challenge of incorporating low frequency items in descriptive grammars of lesser-described languages. The talk examines and contrasts the procedures and steps required when writing a grammar with those necessary to accurately describe a class of expressions that are infrequent and relatively difficult to elicit.

The task of grammar-writing is immense and time and resources one can devote to it are quite often limited. Moreover, the good practice of grammar writing is such that ungrammaticality of certain constructions should be illustrated with examples. All these factors, along with the fact that field linguists quite often set out to document languages they do not know, favours writing descriptive grammars on the basis of elicitation, and on focusing on constructions that can be clearly described on the basis of grammaticality judgements.

However, although elicitation can be effectively used to describe not only morpho-syntax, but also a range of semantic phenomena (Matthewson 2004; Bochnak & Matthewson 2015), it is insufficient for the description of pragmatics or discourse. While this situation can be countered by the use of stimuli and experiments (cf. e.g. San Roque et al. 2012; Silva & AnderBois 2016), these are, at least at their current stage of development, insufficient to document describe the complexities of use of discourse markers. Traditionally, description of discourse marking has been considered rather peripheral to the grammar (cf. e.g. Payne 1997; Chelliah & Reuse 2011). However, in recent years, with a growing interest in interactional linguistics (e.g. Couper-Kuhlen & Selting 2018), interactional phenomena are recognised as being key for adequate and complete description of a language.

In this talk, I discuss this issue on the basis of a particular case study: that of epistemic discourse markers in Upper Napo Kichwa (Quechuan, Ecuador). I show why elicitation and experimental tasks are insufficient for their description, ponder their place in the grammar, and propose a more corpus-grounded approach to descriptive grammar-writing.

References

- Bochnak, M. Ryan & Lisa Matthewson (eds.). 2015. *Methodologies in semantic fieldwork*. Oxford ; New York: OUP.
- Chelliah, Shobhana Lakshmi & Willem Joseph de Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. Dordrecht [Netherlands] ; New York: Springer.
- Couper-Kuhlen, Elizabeth & Margret Selting. 2018. *Interactional Linguistics: An Introduction to Language in Social Interaction*. Cambridge: CUP.
- Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70(4). 369–415.
- Payne, Thomas E. 1997. *Describing Morphosyntax: A Guide for Field Linguists*. Cambridge: CUP.
- San Roque, Lila, Lauren Gawne, Darja Hoenigman, Julia Colleen Miller, Alan Rumsey, Stef Spronck, Alice Carroll & Nicholas Evans. 2012. Getting the Story Straight: Language Fieldwork Using a Narrative Problem-Solving Task. *Language Documentation and Conservation* 6. 135–174.
- Silva, Wilson & Scott AnderBois. 2016. Fieldwork Game Play: Masterminding Evidentiality in Desano. *Language Documentation & Conservation* 10. 58–76.

Language corpora and the syntax-pragmatic interface: the Chipaya subject clitics

Katja Hannß
University of Cologne (Germany)

The aim of the paper is to show the importance of corpus data for grammatical descriptions, particularly at the syntax-pragmatics interface, using the example of the Chipaya language. Chipaya is an endangered isolate of highland Bolivia, spoken by 1,625 people (Cerrón-Palomino 2009:29, 32, referring to Molina & Albó 2006; Adelaar 2007:19). The language has only recently been documented and described (Chipaya documentation project; Cerrón-Palomino 2006). The corpus of 11 hours of speech data on which this study is based results from a DobeS project conducted by the author and colleagues.¹

Among the features in which Chipaya differs from the surrounding Andean languages is a set of subject clitics. These are co-referential with the subject referent of the clause but are not obligatory; i.e. lengthy stretches of discourse can occur without them. In the only substantial grammatical description available so far Cerrón-Palomino (2006:172–173) characterises the Chipaya subject clitics as “floating” and “focalising” their host. Furthermore, Cerrón-Palomino (2006:172–175) claims that the subject clitics can attach to the respective subject pronoun. However, the author’s corpus-based research suggests that things are not as straightforward as has been proposed so far. First, only the subject clitics referring to a first- or second person subject referent are floating and even they are not entirely free with respect to their host but attach either to a constituent indicating the object referent or else to a modifying element (i.e. an adjective or an adverb). They do not occur on the subject noun or pronoun itself. Subject clitics of a third person, in contrast, attach only to the respective subject noun or pronoun. In negative clauses, however, all subject clitics go onto the negation marker. From this results the following syntactic distribution of the Chipaya subject clitics.

	Negation	Subject	Object	Modifying element
1st and 2nd person subject clitics	X	---	X	X
3rd person subject clitic	X	X	---	---

Table 1: Syntactic distribution of the Chipaya subject clitics

It is suggested that from a synchronic perspective the Chipaya subject clitics are indeed focus-marking elements, as claimed by Cerrón-Palomino (2006:172–173), which, however developed diachronically from subject markers. As first- and second persons are more salient, they also make more typical subjects (Siewierska 2004:46) and specific subject markers are thus superfluous. Consequently, the subject clitics of a first- and second person became free to attach to other constituents, thereby developing into focus markers. A third person, however, is less salient and therefore makes a less typical subject (*ibid.*). There, the subject clitics remain *in situ* in order to mark the uncommon third person as subject.

Thus, the syntactic distribution and pragmatic function of the Chipaya subject clitics are more complex than suggested so far. The author’s research therefore demonstrates that corpora of natural language data are essential in analysing the grammar of languages, especially when describing phenomena at the interface of syntax and pragmatics, such as the Chipaya subject clitics. A broad database of natural speech data also enables us to draw conclusions about diachronic processes in languages whose historical development is not well documented.

¹ DobeS = *Dokumentation bedrohter Sprachen* (Documentation of endangered languages).

References

- Adelaar, Willem. 2007. Threatened languages in Hispanic South America. In Matthias Brenzinger (ed.), *Language diversity endangered*, 9–28. Berlin: Mouton de Gruyter.
- Cerrón-Palomino, Rodolfo. 2006. *El chipaya o la lengua de los hombres del agua*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.
- Cerrón-Palomino, Rodolfo. 2009. Chipaya. In Mily Crevels & Pieter Muysken (eds.), *Lenguas de Bolivia. Ámbito andino*, vol. I, 29–77. La Paz, Bolivia: Embajada de Reino de los Países Bajos & MUSEF & Plural Editores.
- Chipaya documentation project. <http://dobes.mpi.nl/projects/chipaya/> (23/09/2018).
- Molina, Ramiro & Xavier Albó. 2006. *Gama étnica y lingüística de la población boliviana*. La Paz, Bolivia: PNUD.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.

On the Role of Natural Discourse Data on Grammar Writing

Yang Huang

Southwest Jiaotong University, China

elvishuang@swjtu.edu.cn

Himmelman (2006) suggested the field of language documentation as ‘concerned with methods, tools...and lasting multipurpose record of a natural language’. Particularly, grammatical features often show different degrees of elaboration in different kinds of discourse (Mithun 2014). In this regard, it is important to draw examples for a grammar not just from monologue, but also from interactive conversation.

This paper examines the basic premise that not only the elicited examples, but also the natural discourse data in speech will provide a snapshot for qualified reference grammar writing. On the basis of the collected data, examples are specifically drawn from the nDrapa language 扎巴语, an endangered Qiangic language of the Tibeto-Burman group spoken in *western Sichuan province* of China. The nDrapa people are the indigenous residents living along the *Yalong River*, who are thought to be connected with the culture of ‘*East Female Country*’ (approx. 6th—7th century A.D) in ancient China.

Firstly, two uncommon grammatical markers are examined to show the significance of interactive speech in reference grammar writing as follows:

ÿ *Humble vs. pejorative modals*

nDrapa uses a preverbal modal marker $nba^{33}ti^{33}$ to encode both humble and pejorative modals. $nba^{33}ti^{33}$ is syntactically attached to nominals rather than predicates. What distinguishes the two functions is based on the distinctive agents in conversation: When the addresser stresses on the self-person expression, $nba^{33}ti^{33}$ works as a humble modal marker (1); When the addresser stresses on the other-person expression, $nba^{33}ti^{33}$ functions as a pejorative modal marker (2).

- (1) $tə^{33}mts^ho$ □□ $ta^{33}ji$ □□ ji^2 □- $kə$ □□- zi □□- $ɣtia^{33}$ - $zɛ^{33}$.
 $nba^{33}ti^{33}$
 LNK then home-HUM DIR-responsible for-PFV.EGO-FACT

After thinking for a while, I determine to raise the humble family (even the family is poor).

- (2) $tə^{33}ta$ □³ po □□ t^ho^{33} - t^ha □□ $ptɛ^ha$ □□ ti^{33} $kə$ □□- hti □³.
 $nba^{33}ti^{33}$
 DEM son-in-law- MOD NUM:CL DIR-put
 PJT

Then, my daughter married a man (son-in-law). It is actually a pejorative

thing.

Historically, a particular morpheme associated with humble modal is apt to derive a pejorative overtone induced by negative context (cf. Grandi & Körtvélyessy 2015). Moreover, the versatile functions of *nba³³tr³³* only appear in discourse and speech genre. In the course of grammar writing, I intentionally concentrate on the positive vs. negative register of the discourse data, refraining from a sentence-by-sentence elicitation, in order to include all the possible contexts where *nba³³tr³³* occurs and how its grammatical character looks like.

ÿ *Thetical*

Secondly, a rich body of examples of theticals will be demonstrated from spontaneous speech in this paper. In nDrapa, a particular postposed demonstrative *mbə³³rə³³* ‘that’ is usually fused with a SAY verb *dε³³* as *dε³³mbə³³rə³³* to signal the extra subjective and personal-judgement modal implication in conversation. Previous studies fail to give a comprehensive account for the grammatical property of this unit. *dε³³mbə³³rə³³* is somewhat similar to the function of a discourse marker. Yet it has more restrictions in terms of the head elements hosted. All the natural discourse data exclusively merit a better understanding on how a thetical intertwines with the discourse context, and how to differ it from default discourse markers (i.e., *me³³ki³³* ‘like that’, *pe³³ma³³la³³* ‘like what you say/hear’) in this language.

Finally, I emphasize that it would be roughly inadequate to take an extreme purist position and exclude elicited data (Pawley 2014). Data collected from natural discourse speech will play a vital role in writing a reference grammar of the minority languages spoken in southwest China.

Abbreviations:

CL: classifier; DEM: demonstrative; DIR: directional marker; EGO: egophoric; FACT: factual evidential; HUM: humble modal; LNK: linker; MOD: mood; NUM: numeral; PFV: perfective; PJT: pejorative modal

References

- Grandi, Nicola & Livia Körtvélyessy. 2015. *Edinburgh Handbook of Evaluative Morphology*. Edinburgh: Edinburgh University Press.
- Himmelman, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Gippert Jost, Nikolaus P. Himmelman, and Ulrike Mosel (eds.), *Essentials of Language Documentation*, pp. 1-30. Berlin & New York: Mouton de Gruyter.
- Huang, Yang. forthcoming. *A Short Grammar of nDrapa*. London: Routledge.
- Mithun, Marianne. 2014. The data and examples: Comprehensiveness, accuracy, and sensitivity. In Toshihide Nakayama & Keren Rice (eds.), *The Art and Practice of Grammar Writing*, pp. 25-52. [Language Documentation & Conservation Special Publication No. 8]. Honolulu: University of Hawaii Press.

Pawley, Andrew. 2014. Grammar writing from a dissertation advisor's perspective. In Toshihide Nakayama & Keren Rice (eds.), *The Art and Practice of Grammar Writing*, pp. 7-24. [Language Documentation & Conservation Special Publication No. 8]. Honolulu: University of Hawaii Press.

Andra Kalnača, Ilze Lokmane
University of Latvia, Riga

**THE LATVIAN DEBITIVE: MORPHOSYNTAX, DISTRIBUTION AND
 PROBLEMS OF DESCRIPTION IN GRAMMAR**

To express the meaning of necessity or obligation in Latvian, a distinct mood form the *debitive* is used (e.g., Mathiassen 1997; Nau 1998; Paegle 2003; Nītiņa, Grigorjevs 2013). The debitive is formed combining the 3rd person present indicative with the prefix *jā-* and the auxiliary *būt* in the finite tense (and mood) form (see also example (2)).

(1) *darīt – dara – ir jā-dara* ‘to do – do – must do’

The prefix *jā-* originates from the genitive singular form of the old Baltic anaphoric and relative pronoun, which is now extinct (see Endzelīns 1981; Holvoet 2001, 2007). There is no corresponding form in other Baltic or Indo-European languages.

From a formal point of view, the debitive construction is unmarked in terms of person and number categories. The subject of the action in the debitive mood is in the dative case (see Fennells 1995), where the dative marks the semantic role of the agent:

(2) *Man ir jā-dara darbs*
 I.DAT be.AUX.PRS DEB-do work.NOM.SG
 ‘I must do the work’ (www.korpuss.lv)

The formation of the debitive mood forms involves a change in the syntactic structure of the clause:

(3) *Indicative*
 a. *Es daru darbu*
 I.NOM do.PRS.1 work.ACC.SG
 ‘I do the work’

as opposed to

Debitive

b. *Man ir jā-dara darbs*
 I.DAT be.AUX.PRS.3 DEB-do work.NOM.SG
 ‘I must do the work’

The debitive, like all other moods in Latvian, has passive forms where the auxiliaries *tikt* ‘to get’ (indefinite forms) and *būt* ‘to be’ (perfect forms) combine with the passive participle which agrees with the semantic object or patient in gender and number, e. g., present indefinite form (more on the typology of passive see, e.g., Klaiman 1991):

(4) *Lūgumam ir jā-tiek pamatotam*
 request.DAT.SG be. AUX.PRS.3 DEB-get corroborate.PTCP.DAT.SG

ar medicīnisko izziņu
 with medical.INS.SG certificate.INS.SG
 ‘The request must be corroborated with a medical certificate’
 (www.mailissima.com)

Due to its morphosyntax and distribution, the description of the debitive in the Latvian grammars has so far been a rather controversial issue (in detail see Nītiņa, Grigorjevs 2013, see also Holvoet, Grzybowska 2014; Lokmane, Kalnača 2014). Endzelīns (1981) argued that the debitive is a variety of the indicative mood, Marvans (1967) and Nau (1998) analyzed it in terms of the so-called modal passive. Andronov (1998) wrote about the *gerundive*, while Holvoet (2001, 2007) suggested a description of the debitive as a particular class of modal forms with indicative, conditional, and oblique forms.

Thus, this paper addresses the following issues:

- 1 The place of the debitive in the mood system of the Latvian language. Is it the verb mood or the passive form, or should it be viewed as a specific group of the verb forms?
- 2 The distribution of the debitive based on the syntactic interpretation of the dative, accusative and nominative case forms of the respective verb, which is a problematic question in the Latvian language.

The examples are extracted from *The Balanced Corpus of Modern Latvian* (available at www.korpuss.lv), and other sources.

References

- Andronovs, Aleksejs V. 1998. „Vajadzības izteiksme” latviešu valodas gramatiskajā tradīcijā. 1. daļa (XVII–XVIII gs.). *Baltu Filoloģija VIII*. Rīga: Latvijas Universitāte, 154–177.
- Endzelīns, Jānis. 1951. *Latviešu valodas gramatika*. Rīga: Latvijas Valsts izdevniecība.
- Fennells, Trevors G. 1995. Teikuma priekšmets latviešu valodas teikumos ar verbu vajadzības izteiksmē. *Lingua Lettica*: 39–46.
- Holvoet, Axel. 2001. *Studies in the Latvian Verb*. Kraków: Wydawnictwo universitetu Jagiellońskiego.
- Holvoet, Axel. 2007. *Mood and modality in Baltic*. Kraków: Wydawnictwo universitetu Jagiellońskiego.
- Holvoet, Axel & Marta Grzybowska 2014. Non-canonical grammatical relations in a modal construction. The Latvian debitive. *Grammatical Relations and their Non-Canonical Encoding in Baltic*. Holvoet, Axel & Nicole Nau (eds.). Amsterdam/Philadelphia: John Benjamins Publishing Company, 97–135.
- Klaiman, Miriam H. 1991. *Grammatical Voice*. *Cambridge Studies in Linguistics*, 59. Cambridge: Cambridge University Press.
- Lokmane, Ilze, Kalnača, Andra. 2014. Modal semantics and morphosyntax of the Latvian DEBITIVE. *Modes of Modality. Modality, typology, and universal grammar*. *Studies in Language Companion Series (SLCS)*. Volume 149. Leiss, E. & W. Abraham (eds.). Amsterdam/Philadelphia: John Benjamins Publishing Company, 2014, 167–192.
- Mathiassen, Terje 1997. *A short grammar of Latvian*. Columbus (Ohio): Slavica Publishers, INC.
- Marvans, Jirži 1967. Par verbu izteiksmes kategoriju mūsdienu latviešu valodā. *Latviešu valodas apcerējumi*. Rīga: Zvaigzne, 127–133.

- Nau, Nicole 1998. *Latvian. languages of the world / Materials 217*. München: LINCOM.
- Nītiņa, Daina, Grigorjevs, Juris (eds.). 2013. *Latviešu valodas gramatika*. Rīga: LU Latviešu valodas institūts.
- Paegle, Dzintra. 2003. *Latviešu literārās valodas morfoloģija*. Rīga: Zinātne.

Variation and language contact in an underdescribed dialect of North Saami

Olle Kejonen, Uppsala University

The Saamic languages and dialects are traditionally described as a linguistic continuum stretching from central Sweden and Norway in the southwest via northern Finland to the Russian Kola peninsula in the northeast. In the seasonal migrations of Saami reindeer herders, national borders have historically been of little importance. Consequently, descriptions of the Saamic varieties made during the 1900s have largely disregarded national borders as linguistic dividing lines or even declared them irrelevant in the description of Saamic dialects (see for instance Nielsen 1926: VIII). In this presentation, I argue that for some syntactic constructions, the Swedish-Norwegian national border does in fact function as a dialectal border as well. I further argue that these isoglosses have emerged as a result of language contact with Norwegian (Germanic, Indo-European) and, to some extent, Meänkieli (Uralic, Finnic). As noted by Ylikoski (2009: 201–202), North Saami is unusual in a European perspective, being a minority language under severe pressure of majority languages from two families unrelated to each other. The analysis presented here is based on preliminary findings from an ongoing project aiming at producing a corpus-based and typologically informed grammar of North Saami, based on the underdescribed and endangered Čohkkiras dialect, spoken in Sweden and Norway.

The formation of polar interrogatives is given special attention as an example of variation within Čohkkiras North Saami. In the eastern subdialect, spoken in Sweden, polar interrogatives are marked simultaneously by constituent fronting and a question marker =*go* ~ =*gos* (1). In the western dialect, spoken in Norway, polar questions can be formed with mere constituent fronting (2).

- | | | | | |
|-----|---------------------------|-------------|-----------------|----------------|
| (1) | <i>osttiide=go</i> | <i>doai</i> | <i>mielkki?</i> | |
| | buy.2DU.PST=Q | 2DU.NOM | milk.ACC.SG | |
| | ‘Did you (two) buy milk?’ | | | |
| (2) | <i>lehpēt</i> | <i>dii</i> | <i>oastán</i> | <i>biepmu?</i> |
| | be.2PL.PRS | 2PL.NOM | buy.PRF | food.ACC.SG |
| | ‘Have you bought food?’ | | | |

I argue that the western construction is a result of intense contact with Norwegian, and I also make some comparisons with other Saamic languages. Paradoxically, in drawing closer to the majority language Norwegian, the western subdialect favors a typologically unusual way of marking polar questions, almost exclusively attested in Western Europe (Dryer 2013).

Variation and language contact in an underdescribed dialect of North Saami

Olle Kejonen, Uppsala University

References

Dryer, Matthew S. (2013). *Polar Questions*. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/116>, Accessed on 2018-09-28.)

Nielsen, Konrad (1926). *Lærebok i lappisk: utarbeidet på grunnlag av dialektene i Polmak, Karasjok og Kautokeino*. Oslo: Brøgger

Ylikoski, Jussi (2009). *Non-finites in North Saami*. Helsinki: Société Finno-ougrienne

Grammaticality judgements in language contact: the case of Mano, Mande

This paper focuses on the challenges that a fieldworker and grammar writer faces in a language contact situation, taking as an example the reflexive construction in the Mano language, spoken by 400 000 speakers in Guinea and Liberia. Mano has a dedicated 3rd person sg. reflexive pronoun *ē* which is in contrast with the basic 3rd person sg. pronoun *à*. In all other persons and numbers there is no dedicated reflexive pronoun. These pronouns are used in direct object, argument of postposition and inalienable possessor position. Typically, the reflexive pronoun has a locally bound reading (has to be coreferential to the subject of the first finite clause), while the basic *à* pronoun is locally free (ex. 1, 2).

Mano has long been in intense contact with Kpelle and manifests a number of contact-induced phenomena in phonology and morphosyntax (Khachaturyan 2018). Many Mano speakers speak Kpelle as an L2 or learn it in childhood being raised in bilingual families. Crucially, the Kpelle language lacks a dedicated reflexive pronoun, using a 3rd person sg. pronominal prefix for both bound and free readings (3). In contrast with native Mano adults, bilingual Mano-Kpelle children can, along with the reflexive pronoun, use the basic pronoun in relevant contexts (ex. 4, recorded from a 10 y.o). Similarly, native speakers of Kpelle who learned Mano as adults rarely use the reflexive pronoun and prefer the basic pronoun.

Occasionally, in natural speech of native Mano speakers the basic pronoun in some peripheral syntactic positions can be used to refer to the local subject (ex. 5). In grammaticality judgements, when Mano speakers are asked to provide “acceptable”, but not only “good” examples, examples with a basic pronoun in a bound interpretation are marginally accepted (ex. 6).

There is a context where variation between a reflexive and a basic pronoun is systematic: it is the case of the inclusory construction, where the first element, an inclusory pronoun, refers to the entire conjoined set, and the noun phrase following it refers to an included subset. The variation attested in native texts is about 50% reflexive pronoun to 50% basic pronoun. 7a with a reflexive pronoun has only a bound reading, while 7b with a basic pronoun has both a bound and free readings.

In a different paper, I argued that the inclusory construction was borrowed from Kpelle into Mano (Khachaturyan 2019). The equivalent construction in Kpelle is given in 8. The reason why in the Mano construction both basic and reflexive pronouns are possible is probably because in the source Kpelle construction there was no reflexive pronoun, so as a result of borrowing and continuous exposure to the source language, two variants became possible: one representing a direct calque of the source construction (with a basic pronoun) and another adjusted to the Mano morphology (with a reflexive pronoun). The examples with the basic pronoun like 7b were obtained in natural speech of native adult Mano speakers.

The performance of L2 speakers in a given language may differ dramatically from the performance of L1 speakers because of interference phenomena (compare 1 and 2 vs 4; cf. Demirci 2001). At the same time, L2 performance may influence the grammaticality judgements of native speakers of the language in question – they have heard some of those “erroneous” examples, so they become part of what can be said and understood (ex. 6). In the long run, contact between the two languages may produce stable effects in the grammars (ex. 7). The main difficulty, then, is to assess the scope of variation at a given stage and judge what belongs to the grammar of one language and what is arguably foreign. What decision should a grammar writer take regarding example 5: is it a legitimate example of Mano? Is it a slip of a tongue, contact-induced influence or should it be considered part of the acceptable, although marginal, variation?

References

- Khachaturyan, Maria. 2018. A sketch of dialectal variation in Mano. *Mandenkan* 59, 31-56.
 Khachaturyan, Maria. 2019. Inclusory constructions in Mande: the emergence of a typological rarum. To appear in *Folia Linguistica*.

Konoshenko, Maria. 2017. Кпелле язык [The Kpelle language]. In Valentin Vydrin, Julia Mazurova, Andrey Kibrik & Elena Markus (eds.), *Языки мира. Языки манде* [Languages of the world: Mande languages], 284–343. Saint Petersburg: Nestor-Istorija.

Demirci, Mahide. 2001. Acquisition of binding of English reflexives by Turkish L2 learners: A Neo-Gricean pragmatic account. *Journal of Pragmatics* 33. 753-775.

Examples

1a $l\bar{e}$ \bar{e} $k\grave{o}$ $g\grave{e}$ - $p\grave{e}$ $\grave{l}\grave{e}$ vs 1b $l\bar{e}$ \grave{a} $k\grave{o}$ $g\grave{e}$ - $p\grave{e}$ $\grave{l}\grave{e}$
 3SG.EXI 3SG.REFL hand see-INF 3SG.EXI 3SG hand see-INF
 'She is looking at her own hand' vs 'She is looking at his hand' (elicited)

2b \bar{e} $b\grave{a}$ \bar{e} $m\grave{o}$ vs 2b \bar{e} $b\grave{a}$ \grave{a} $m\grave{o}$
 3SG.PST touch 3SG.REFL PP 3SG.PST go 3SG PP
 'He touched himself' vs 'He touched her'. (elicited)

3 \check{e} $l\acute{o}$ $j\grave{e}$ i $b\grave{e}$ \acute{e} \grave{r}
 3SG.CONJ enter 3SG\hand.LOC DEF\house.LOC
 'He entered his own/her house.' (Konoshenko 2017: 299; p.c.)

4 \bar{a} \grave{a} $p\acute{e}$ \grave{a} $k\grave{o}$ $d\bar{o}$ \grave{a} $m\grave{o}$, \grave{a} $s\acute{e}$ \acute{r} $m\grave{o}$
 3SG.PRF just 3SG hand put 3SG on 3SG side on
 'He put his hand on himself, on his side.' (example from Family Problem Picture Task, natural)

5 \bar{e} $n\bar{u}$ \grave{a} $p\grave{a}$.
 3SG.PST come 3SG.NSBJ chez
 'He came back home (lit.: he came at him).' (natural)

6 $l\bar{e}$ \grave{a} $z\acute{u}$ \acute{u} - $p\grave{e}$ $\grave{l}\grave{e}$
 3SG.EXI 3SG wash-INF
 ??? 'She is washing herself' (elicited)

7a $w\grave{a}$ \bar{e} $l\grave{o}$ $k\acute{o}$ \acute{o} 7b $w\grave{a}$ \grave{a} $l\grave{o}$ $k\acute{o}$ \acute{o}
 3PL.IP 3SG.REFL mother 3PL.IP 3SG mother
 'he and his_{i,j} mother (lit.: they his own mother)' vs 'he and his_{i,j} mother (lit.: they his mother)' (elicited)

Enets clause-combining: how it really works¹

Olesya Khanina¹, Maria Ovsjannikova^{1,2} & Andrey Shluinsky¹

(1: Institute of Linguistics, Russian Academy of Sciences; 2: Institute for Linguistic Studies, Russian Academy of Sciences)

This paper is devoted to clause combining in the two dialects of Enets, Forest and Tundra. In this Uralic, Northern Samoyedic language spoken in the North of central Siberia, clause combining is mainly realized by the use of non-finite verb forms; the uses of sentence-like strategies are attested, but most of them are clearly influenced by Russian, with the exception of complements of mental verbs and verbs of speech (~ direct speech strategy). Nominally, the system of non-finite forms is significantly richer in Forest Enets (13 forms) than in Tundra Enets (9 forms), as the former has two more participles and two more converbs; however, the three of these FE-only non-finite forms are used very rarely, so this difference may be an artefact of our description, as our Forest Enets corpus is three times bigger than our Tundra Enets corpus (21 hours / 30 000 sentences for FE; 7 hours / 10 000 sentences for TE).

Table 1 gives a summary of all Enets non-finite forms. Some observations over the system can be the following. First, most of the Enets non-finite forms can be used both for same-subject and different-subject clause-combining, with clear frequency preferences of most forms for same-subject uses. Second, non-finite forms with the highest token frequencies (CVB, PTCP.SIM, CVB.COND, NMLZ) are multifunctional in the sense that they are attested in at least two types of dependent clauses: adverbial, relative, or complement. Third, adverbial and complement clauses share most strategies, while relative clauses often use a dedicated verb form. Forth, there are no non-finite forms that would be used exclusively in complementation, while a significant share of non-finite forms is used either in relative or in adverbial clauses only.

We aim to analyze Enets clause combining as a system: matching all available non-finite forms to all possible types of dependent clauses, with particular attention to the ways of expression of the core arguments in the dependent clause. Besides, we comment on actual frequency of specific constellation of parameters in the Enets corpora, so that a static description of what is possible in the language is actually complemented by a dynamic description of how the system is actually used by the speakers. Finally, we track some changes at the microdiachronic level, contrasting the uses of the non-finite forms in the speech of the last two generations of the Enets speakers. Thus, this paper contributes to several themes of the conference: description of specific grammatical structures in an underdescribed language, ways to embed language variation in a grammatical description (on dialectal and intergenerational levels), and the use of corpora in grammar writing allowing for frequency and usage related observations.

¹ This study was supported by a grant 17-34-01068-OGN from the Russian Foundation for Basic Research.

Table 1. Enets non-finite forms and their functions

Gloss	Main functions	Number of tokens in the Forest Enets corpus	Used in a different-subject clause	Complement clauses	Relative clauses	Adverbial clauses
CVB	Adverbial clauses of purpose, manner, anteriority, and simultaneity; complement clauses of phasal, modal, emotive, mental, and evaluational predicates; marginal relative clauses	1606	rarely	+	rarely	+
PTCP.SIM	Simultaneous relative clauses; adverbial clauses of simultaneity (with Dative)	1093	+	-	+	+
CVB.COND	Adverbial clauses of condition	870	+	+	-	+
PTCP.ANT	Adverbial clauses of anteriority	610	+	-	+	-
NMLZ	Complement clauses of phasal, modal, emotive, mental, and evaluational predicates; adverbial clauses of simultaneity and purpose (with postpositions)	413	+	+	-	+
NMLZ+ABL	Adverbial clauses of anteriority	275	+	-	-	+
CVB.SIM	Adverbial clauses of simultaneity	187	+	-	-	+
PTCP.ANT.PASS (FE only)	Anterior relative clauses	109	+	-	+	-
SUP	Adverbial clauses of purpose (only with verbs of movement)	37	-	-	-	+
PTCP.POST (FE only)	Posterior relative clauses	32	+	-	+	-
PTCP.ANT.NEG	Anterior relative clauses in case the event has not yet taken place	11	+	-	+	-
CVB.ANT (FE only)	Adverbial clauses of anteriority	8	-	-	-	+
JUSS.NEG (FE only)	Adverbial clauses of negative purpose	5	+	-	-	+

Challenges of an Athabaskan grammar (Andrej Kibrik, Institute of Linguistics, RAS)

Athabaskan language family, the core of a larger Na-Dene family, is one of the major genealogical groups of North America. Edward Sapir (1929) used to say that the Athabaskan languages are perhaps the most “specialized” among the native languages of North America. Athabaskan languages are also quite distinct against the background of world-wide linguistic diversity. Some of the typological hallmarks of Athabaskan include:

1. Qualitative morphological complexity. Many native American languages are complex in a quantitative way (polysynthesis). But Athabaskan languages, on top of that, are also unusually complex in a qualitative way: the meaning-to-form relationships are often opaque; there is extensive homophony of morphemes; much fusion, allomorphy and sensitivity to inflectional classes; extremely complex morphophonemics.
2. Prefixation. Athabaskan languages are probably the most prefixing languages in the world. There are other exclusively prefixing languages, but they are only moderately synthetic. Suffixation, which is a cross-linguistic default, is minimal in Athabaskan and, synchronically, has mostly developed into root allomorphy.
3. Violation of scope. Contrary to cross-linguistic tendencies, Athabaskan verb template demonstrates an unusual mix of inflectional and derivational positions. Most of inflection is closer to the root than most of derivation (cf. Rice 2000).
4. Extreme concern for expressing aspectual meanings. There are several zones in the morphological structure where various kinds of aspect-related meanings can be conveyed, including choice of root variant, derivational and inflectional prefixes (cf. Kari 1979). In certain word forms up to six aspect-related morphemes may be found.
5. Unusual patterns of event conceptualization. In Athabaskan languages, lexical content of the verb root is much less specified than in most other languages (Author 2012). On the other hand, in some domains verb semantics is very specific, including class affiliation of the absolutive argument (Author 2017).

These and other typological peculiarities definitely deserve interest of typologists. However, there are few Athabaskan grammars, and in those that are available linguistic structures are often described in a less than transparent way, partly because of the traditional idiosyncratic terminology.

In this paper I am going to report my current project of an Upper Kuskokwim grammar. Upper Kuskokwim is a moribund language of interior Alaska, only spoken by a handful of older persons. Upper Kuskokwim is among the most conservative Athabaskan languages and is thus a good representative of the family as a whole. While Athabaskan languages of Alaska are fairly well documented with respect to lexicon (e.g. Jetté and Jones 2000), there is no full grammatical description for any of those languages. I am preparing a typologically oriented grammar of Upper Kuskokwim, intended to make the evidence of this most interesting language family accessible to a wide range of linguists. The manuscript is currently over 1000 pages long. It includes information on all aspects of the language, including history, ethnography, sociolinguistics, dialectal and idiolectal variation, segmental phonetics, prosody, morphophonemics, morpheme and word classes and structure, all lexical categories and their structures, lexical and grammatical semantics, clause-internal syntax, clause combinations, local and global discourse structure, etc. The data comes from both elicitation and natural discourse. Along with the grammar project, there is a parallel project of language documentation, based on a corpus of discourses representing a variety of genres. This work is done in contact with the native communities of the villages of Nikolai and Telida where the language is/was spoken. In addition to sharing this work with the conference participants, I hope to gain much from the experience of other linguists involved in grammar writing.

Jetté, Jules & Eliza Jones. (2000) *Koyukon Athabaskan Dictionary*. Edited by James Kari. Fairbanks: Alaska Native Language Center.

Kari, James. (1979) *Athabaskan verb theme categories: Ahtna*. Fairbanks: Alaska Native Language Center.

Rice, Keren. (2000) *Morpheme order and semantic scope: Word formation in the Athapaskan verb*. Cambridge: Cambridge University Press.

Sapir, Edward. (1929) Central and North American languages. *Encyclopedia Britannica*, vol. 5, 138-141.

Author (2012)

Author (2017)

Voice alternations and non-actor orientedness in comitative verbs in Northern Luzon

Sergei Klimenko (Institute for Linguistic Studies, Russian Academy of Sciences)

Comitative verbs with the prefix *maki-/paki-* (or its cognates) are well known in Philippine languages. Two phenomena related to such verbs, however, have gone largely unnoticed: at least some Northern Luzon languages have non-actor voice and non-actor oriented comitative verbs.

Such forms first drew my attention during my fieldwork on the previously undocumented Yattuka language, where actor voice comitative verbs have counterparts in other voices: e.g., *moki-inum* (*p<um>oqi-ʔ~ʔinum*) ‘to drink with somebody’ with the accompanee in the subject position vs. *pokiinnuman* (*poqi-ʔi<n~>num-an*) with the companion in the subject position. Later, I found that non-actor voice comitative verbs also exist in at least seven other Northern Luzon languages. However, they either can be located only in examples in dictionaries or bible translations (Keley-i (Hohulin et al. 2018), Amganad Ifugao (Bible), Northern Kankanay (Wallace 2018)) with no mention in the available grammar descriptions (Hohulin et al. 2018; Cunningham 1965; West 1973; Sawyer 1975; Porter 1979), or they are dealt with in a grammar description or dictionary entry in a range of various terms other than voice (transitive Undergoer collective verbs in Ibaloy (Ruffolo 2004: 253), object cross-referencing in participatory verbs in Tuwali Ifugao (Hohulin and Hohulin 2014: 46, 502), genitive concomitant aspect verbs in Batad Ifugao (Newell and Poligon 1993), nominalization of social verbs in Ilokano (Rubino 2000: 416). Such items are verbs, rather than nominalizations, as they manifest both voice (often referred to as ‘focus’ in Philippine linguistics) and aspect morphology, which is commonly used as a basis to classify lexemes as verbal in Philippine languages (cf., for instance, “Verbs in Ilocano can be distinguished from other lexical classes because they are marked for Aspect ... and Focus/Transitivity...” (Rubino 1997: 186)).

The lack of description and this terminological variety might be the reason why non-actor voice comitative verbs have never been attested for Philippine languages in typological works. Liao in her study of Philippine comitative verbs claims that forms with *paki-* are gerunds, nominalizations, imperatives, dependent forms, imperfectives or requestives and explicitly refutes the possibility that such verbs have voice paradigms (2011: 212-214). Stolz et al. suggest that Tagalog actor voice comitative verbs have non-actor voice counterparts marked with the prefix *ka-*, e.g. *maki-pag-ʔusap* ‘AV.COM-STEM-converse[NEUT]’ vs. *ka-ʔusap-in* ‘CA-converse-PV[NEUT]’ (2006: 59). However, verbs with *ka-* do not comprise a single voice paradigm with verbs with *maki-* derived from the same root for two reasons. First, many verbs with *ka-* are not comitative, since the participant in the subject position with them does not necessarily perform the same action (e.g., the non-actor participant of *ka-ʔusap-in* can be asleep). Second, verbs with *ka-* have actor voice counterparts with the same prefix (e.g., *k<um>a-ʔusap* ‘<AV>CA-converse[NEUT]’). Voice paradigms in comitative verbs are also highly unusual globally, as they are not attested either in another broad typological study (Arkhipov 2005, 2009), and while the accompanee normally holds a higher syntactic status (Arkhipov 2005: 75), in non-actor voice comitative constructions the ranking switches.

Non-actor oriented comitative verbs (i.e. with non-actor pluralized participant) have never been attested in Philippine languages to my knowledge, but they can be found in dictionary examples and bible translations of at least four of the mentioned languages (Keley-i, Tuwali Ifugao, Batad Ifugao, and Amganad Ifugao), and the field data of Yattuka. Below is a Batad Ifugao example of a construction with a patient voice comitative verb, whose pluralized participant is the patient:

Niʔ-patoy	Alanghabon	nan	hamuti=hnan	doʔʔo.
PFV.COM-kill[PV]	PN.Alanghabon	NOM.MED	bird=OBL.MED	rat

‘Alanghabon included the bird with the rat in killing.’ (Newell and Poligon 1993)

This study presents the data on the two understudied phenomena in question and discusses how it is interpreted in the existing descriptions of the said languages.

Abbreviations:

AV, actor voice; CA, counter-agent; COM, comitative; MED, medial demonstrative; NEUT, neutral aspect; NOM, nominative; OBL, oblique; PFV, perfective; PN, personal noun; PV, patient voice; STEM, stem-deriving prefix.

References:

- Arkhipov, Alexandre V. 2005. *Typology of Comitative Constructions* (In Russian: *Типология комитативных конструкций*). PhD thesis. Moscow State University.
- Arkhipov, Alexandre V. 2009. Comitative as a cross-linguistically valid category. In *New Challenges in Typology. Transcending the Borders and Refining the Distinctions*, ed. by Patience Epps and Alexandre Arkhipov. Berlin, New York: Mouton de Gruyter. Pp. 223-246.
- Bible. Retrieved September 27 2018, from <http://listen.bible.is/IFAWBT/Gen/1>.
- Cunningham, Margaret C. 1965. *Another sketch of Ifugao grammar*. Manila: SIL.
- Hohulin, Richard M. and B. Lou Hohulin. 2014. *Tuwali Ifugao Dictionary and Grammar Sketch*. Linguistic Society of the Philippines.
- Hohulin, Richard M., B. Lou Hohulin, Alberto K. Maddawat. 2018. *Keley-i Dictionary and Grammar Sketch*. SIL International. Retrieved September 27 2018, from <https://keley-i.webonary.org>.
- Liao, Hsiu-chuan. 2011. On the Development of Comitative Verbs in Philippine Languages. In *Language and Linguistics* 12.1: 205-237.
- Newell, Leonard E. and Francis Bon'og Poligon. 1993. *Batad Ifugao dictionary, with ethnographic notes*. Linguistic Society of the Philippines. Retrieved September 27 2018, from https://philippines.sil.org/resources/online_resources/ifb.
- Porter, Doris. 1979. Northern Kankanay morphology. In *Studies in Philippine Linguistics* 3.2: 20-62.
- Rubino, Carl Ralph Galvez. 1997. *A Reference Grammar of Ilocano*. PhD thesis. University of California, Santa Barbara.
- Rubino, Carl Ralph Galvez. 2000. *Ilocano dictionary and grammar: Ilocano-English, English-Ilocano*. University of Hawai'i Press books.
- Ruffolo, Roberta. 2004. *Topics in the Morpho-Syntax of Ibaloy, Northern Philippines*. PhD thesis. The Australian National University.
- Sawyer, Louise. 1975. Aspect in Amganad Ifugao. In *Anthropological Linguistics* 17.3: 107-116.
- Stolz, Thomas, Cornelia Stroh, and Aina Urdze. 2006. *On Comitatives and Related Categories: A Typological Study with Special Focus on the Languages of Europe*. Berlin: Mouton de Gruyter.
- Wallace, Judy (ed.). *Northern Kankanay - English Dictionary*. SIL International. Retrieved September 27 2018, from northernkankanay.webonary.org.
- West, Anne. 1973. The semantics of focus in Amganad Ifugao. In *Linguistics* 11: 98-121.

The challenges and advantages of representing variation in grammars

Martin Kohlberger (Leiden University/University of Texas at Austin)

Traditionally, the main purpose of a grammar is to serve as a catalogue of the linguistic structures found in a particular language. However, describing and analysing a grammatical structure can be difficult if there is variation in the data. One of the solutions is to consider some of that variation to be irrelevant to the larger analysis and to abstract away from it. For example, much of the phonetic and phonological variation in natural speech is easily smoothed over during the transcription process in which continuous speech is transformed into a discrete orthographic sequence. Other forms of variation, such as code-switching or multilingual discourse, might be actively avoided by a grammar writer, in order not to complicate the description of one language with elements from another.

An alternative approach is to attempt to convey as much of the variation as possible in the grammar. The advantage of this approach is that the resulting analysis is more representative of the data that it is based on. However, this approach comes with methodological difficulties. Clear-cut generalisations about grammatical structures might be harder to make. And there may be practical typographical hurdles to conveying variation. For example, what should the citation form of a morpheme be if it has multiple different versions in the data: the most frequent form, the most archaic form (if it is possible to determine this) or a list of all the forms? Should morphemes which exhibit variation be glossed using an abstract underlying form or should the glosses also represent the variation in the data?

In this talk, the challenges of representing variation in a grammar will be discussed, based on the author's experience of writing a grammar of Shiwiar, a previously undocumented Chicham (Jivaroan) language spoken in the lowlands of eastern Ecuador and northern Peru. Not only does Shiwiar exhibit complex morphophonological processes that result in rich allomorphic alternations, but there is substantial interspeaker variation found in all areas of Shiwiar grammar. Additionally, Shiwiar speakers are mostly trilingual, and they often code-switch into Spanish.

Although conveying such a degree of variation was difficult from an organisational and typographic perspective, the resulting grammatical description automatically highlighted areas of Shiwiar grammar which are currently undergoing change. In this sense, a grammatical description which acknowledges variation is not only useful as a catalogue of linguistic structures that are currently present in a language, but also as a descriptor of language change in progress.

Elder South Saami speaker's innovative language vs. prescriptive language: The impact on language description.

Richard Kowalik, PhD student, Department of linguistics, Stockholm university

This talk addresses the complex of problems of different languages in use among different generations of South Saami speakers with respect to describing the language's grammar.

South Saami (ISO-code *sma*) is the southwestern-most language of the Saamic sub-branch of Finno-Ugric languages, spoken in Sweden and Norway. It is a small, endangered language with approximately 300–500 speakers, consisting of a small group of native speakers (L1-speakers) among the oldest generation, and a growing number of New Speakers (L2-speakers) among the younger generation.

The South Saami spoken by the older generation could be said to reflect the most innovative status of the language, which contrasts with the revitalized language spoken by younger generations. South Saami has an official orthography since 1978, it is taught in school in different places, and it is undergoing processes close to standardization. Not surprisingly, the teaching is based on a prescriptive use of language, and often more conservative. The differences between this innovative language of L1-speakers and the prescriptive use of language of L2-speakers extend into most grammatical domains, such as phonology, morphology and morphosyntactic features, into basic word order and different attitudes towards loan words and calques.

In this talk, I will contrast the use of pro-forms by older native speakers with prescriptive language use. I will argue for an expansion in function, or possibly ongoing processes of grammaticalization, of forms in the language by older speakers. I support my view based on data that I am collecting in collaboration with first language speakers. Most pro-forms can be inflected for case, number and person (see Appendix Table 1a on page two). However, some morphological forms do not occur in my data of natural, free speech – not only because they are low frequent, but because the function of one form is expanded and used in other contexts as well. In example (1), taken from the SIKOR corpus and reflecting prescriptive language use, the relative pronoun (*mah*) agrees in number with the subject; in example (2), taken from free speech recorded by the author in 2017, we find a plural subject but the relative pronoun (*mij*) in the singular:

- (1) *Dah guaktah fenomeen-h mah Haspelmath*
 3PL two.NOM.PL phenomenon-NOM.PL REL.NOM.PL Haspelmath

gohtje equipolleente jih antikausatijve [...]
 name.PRS.3SG equipollent and anti-causative

‘These two phenomena, which Haspelmath calls equipollent and anti-causative [...]
 [source: SIKOR]

- (2) *seamma goh dah moderne ‘transitspår’ mij*
 same as 3PL modern transit.track.PL REL.NOM.SG

daelie Bienjedaelie-n bijre gååvnesieh
 now Funäsdalen-GEN.SG around exist.PRS.3PL

‘the same as these modern ‘transit skiing tracks’, which now exist around Funäsdalen.’
 [Own data: sma20170922i]

The results are placed both within a typological context of the surrounding languages, and within general processes and tendencies of grammaticalization and language change. I will point out challenges when writing a grammar for the language and discuss my solutions to it.

Appendix

Table 1a: Example paradigm: The paradigm for the relative/interrogative pronoun *mij*, according to prescriptive language (Magga & Magga 2012:56)

<u>Case</u>	<u>SG</u>	<u>PL</u>
NOM	mij	mah
GEN	man	mejtie
ACC	maam	maj
ILL	misse	mejtie
LOC	mesnie	mejnie
ELA	mestie	mejstie
COM	mejnie	mejgumie

Table 1b: The relative/interrogative pronoun *mij* and its forms attested in the author's data. Forms in parentheses occur only marginally (once or twice), or in elicitation context.

<u>Case</u>	<u>SG</u>	<u>PL</u>
NOM	mij	(mah; meh)
GEN	man	-
ACC	maam	-
ILL	(misse)	-
LOC	-	-
ELA	-	-
COM	(mejnie)	-

References

Magga, Ole Henrik, and Lajla Mattsson Magga. 2012. *Sørsamisk grammatikk*. Kárásjohkka: Davvi girji.

SIKOR. UiT The Arctic University of Norway and the Norwegian Saami Parliament's Saami text collection, *Version 08.12.2016*, URL: <http://gtweb.uit.no/korp/>

Demonstratives in Mozambican Ngoni

Heidrun Kröger

Goethe University Frankfurt/Main, SIL International

heidrun_kroeger@sil.org

The proposed paper presents data on the demonstrative system of a hitherto largely undescribed Bantu language, namely Mozambican Ngoni.

“In demonstratives, many Bantu languages have a three-way contrast (‘near speaker, near addressee / previously referred to, far from both’), as Nurse and Philippson (2003: 9) point out. When one looks through Ngoni texts it becomes obvious that Ngoni has more demonstrative forms than this and in general makes much use of demonstratives. In sorting through the different forms to discover their functions, I found Diessel’s study (1999) on demonstratives from a typological viewpoint an immense help.

As it turns out, Mozambican Ngoni has a person-oriented system with two additional factors that are morphologically distinguished: physical contact and visibility. Hence, the distinction is fivefold:

- near the speaker with physical contact,
- general near the speaker,
- near the hearer,
- far from speaker and hearer, but within view,
- general far.

Apart from these forms there is a separate identificational proximal demonstrative, as well as emphatic forms. These forms are briefly introduced according to, their morphological make-up, their semantic features and their position in the noun phrase. It is shown that the main position of the demonstrative is phrase-final, but remarkably, demonstratives can occur pre- and post-nominally as well, and may appear up to three times within the same noun phrase.

As for the usage in texts, two of the five basic demonstratives, the general distal and the medial are used endophorically to indicate definiteness; the distal in the case of anaphoric reference, the medial in the case of associative anaphora, establishing relative clauses and other situations of definiteness in the immediate discourse.

Diessel, Holger. 1999. *Demonstratives. Form, Function and Grammaticalization*. [Typological Studies in Language 42]. Amsterdam: John Benjamins.

Kröger, Heidrun. 2008. *Algumas Notas Gramaticais sobre Xingoni*. Nampula: SIL. Available at <http://lidemo.net/>. 2017 update available on Researchgate.

Kröger, Oliver. 2014. *Personenreferenz in Erzähltexten des mosambikanischen Ngoni*. Dissertation: Köln.

Nurse, Derek & Gérard Philippson (eds.). 2003. *The Bantu Languages*. London & New York: Routledge.

African Basic Grammar. Review and perspectives after 12 years

Oliver Kröger (SIL International)

This paper presents the principles that underlie the methodology called the African Basic Grammar (ABG) and provides a place for critical review and feedback in the light of language development experiences.

ABG was designed to produce entry-level grammatical sketches of underdescribed African languages. Its three components are: a) Typological parameters for parts of speech b) ready-made lectures and c) shell books for write-ups.

The typological parameters are taken from Heine and Vossen (1981). For any given linguistic variety, three questions are asked: Do nouns fall into formal categories, like noun-classes or grammatical gender? Are there productive processes of verbal derivation, which change a basic verbal stem into a derived form with systematic changes in meaning and function, as in causative, applicative or passive verbs? Do nouns show formal indication of nominal case, as in genitive, ablative, possessive? Since the answers are all binary, eight basic language types result.

The lectures were developed for mother-tongue speakers and others interested in language development. First comes the core idea, that serves as a didactic guideline: For example, in a type IV language like Balante (West-Atlantic), the elaboration of data will centre around the nominal classification into 11 singular-plural nouns, whereas neither nominal case marking on the noun nor verbal extension need any attention. In a type VII language like Bertha (Nilo-Saharan isolate) only the four or more cases signalled by nominal prefixes need discussion. During a given workshop, the set of 10 lectures varies according to the inventory of each language.

The shell-books provide templates to enter language data during a workshop: A page on the language situation is followed by an overview of the orthographic system(s) used, then the main sections contain every part of speech, each with a brief and informal definition and ready-made illustrations, allowing speakers without formal training to grasp the basic ideas needed for language development activities (lexicography, translation, writing, standardization).

The publication also contains a small sample text and an overview of publications about the language in question. Often, the grammar sketch that speakers produce in a Discovery workshop is the first ever written material about a spoken language that is available to the ethnolinguistic community and a national audience.

What can be achieved in a two-weeks workshop in terms of valid data collection and to what extent does a typological approach contribute to language development? What are potential benefits for further typological research?

Heine, B. & R. Vossen 1981. Sprachtypologie, in: B. Heine, Th.C. Schadeberg & E. Wolff (eds.): Die Sprachen Afrikas, pp. 407-444. Hamburg: Helmut Buske.

Kröger, O. 2008. Bridging the gap: Bantu grammar workshops as a means to support community-based language development, in: *Anuário Internacional de Comunicação Lusófona 2008: 157-180*. LUSOCOM Federação Lusófona de Ciências da Comunicação.

Kutsch-Lojenga, C. 1996. "Participatory research in linguistics." in: *Notes on Linguistics* 73: 13-27.

Vella, J. 1995. Training through dialogue. Promoting effective learning and change with adults. The Jossey-Bass higher and adult education series.

Describing non-basic constituent order in Palaung (Austroasiatic)

Wei-Wei Lee & Rachel Weymuth - University of Zurich, Department of Comparative Linguistics

One aspect in which the world's languages show considerable diversity is the way in which they order elements in a sentence. As the WALS (Dryer 2013) shows, some possible constituent orders are far more frequent than others. As valuable as this resource is for linguistic typologists, it inevitably copes with problems regarding sampling and the granularity of language descriptions. This becomes clear when we take a closer look at the situation in Southeast Asia, where WALS displays a distorted picture of reality. It will be shown that the exclusive focus on basic word order and the lack of descriptions of smaller, less prestigious languages overshadow many observations that are likely to have important implications for our understanding of word order typology and the history of the languages in the area. It is thus evident that descriptions of lesser known, underdescribed languages looking beyond the dominant constituent order in main clauses are much needed.

We will present preliminary findings from two underdescribed Austroasiatic language varieties, Palaung Rumai (*rbb*) and Palaung Ruching (*pce*). The varieties belong to the Palaungic branch and are spoken in Myanmar. The data, including elicited and natural speech, were independently collected by the present authors in Mandalay Region and Shan State. Rumai and Ruching display basic SVO order in main clauses, in agreement with the general pattern in Mainland Southeast Asia. In certain types of subordinate clauses and other specific environments, however, VSO order is employed. This is illustrated below by examples containing a temporal (1) and a relative (2) clause.

(1) Rumai (Palaungic, Austroasiatic)

br: hōm k̄: t̄-pōm s̄l̄w k̄t̄â:j d̄i: gi:j-t̄v̄h k̄:
when eat 3PL OBL1-rice dog ground MEDL PROG-look 3PL
 'When they were eating, the fox was watching them.'

(2) Ruching (Palaungic, Austroasiatic)

bap hliu ?əu di m̄ hma:ɔ m̄ bleh
book give 1SG GOAL 2SG where 2SG put
 'Where did you put the book I gave you?'

From both an areal and genealogical perspective, the occurrence of a verb-initial structure is unexpected. As the order occurs in a more peripheral part of grammar, large-scale comparisons like WALS would not have included this feature. However, this finding is just as relevant for synchronic and diachronic typologies of syntax. Closer examination of other Austroasiatic languages has, for instance, shown that verb-initial order is more frequent than previously assumed (Jenny *forthc.*). This bears strong significance for the reconstruction of Proto-Austroasiatic syntax, as history does not typically manifest itself in a limited area of grammar only. With our descriptive work on Palaung varieties, we aim to contribute to a more comprehensive understanding of the cross-linguistic diversity of constituent order.

References

- Dryer, Matthew S. 2013. Order of Subject, Object and Verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Jenny, Mathias. *Forthcoming*. Verb-initial structures in Austroasiatic languages.

Describing Unserdeutsch (Rabaul Creole German): Methodological and empirical challenges

*Péter Maitz & Siegwalt Lindenfelser
(University of Bern)*

Since 2014, an international team of researchers has been working on the documentation of Unserdeutsch, a still widely unknown and critically endangered German-based creole language of Papua New Guinea (cf. Maitz & Volker 2017, Götze et al. 2017, Lindenfelser & Maitz 2017). After extensive fieldwork data collection in the Pacific and the development of an annotated digital corpus, the next stage of the project is supposed to be the systematic corpus-based description of the phonological and grammatical structure of Unserdeutsch. In our paper we will address some of the major methodological and empirical challenges of this enterprise.

- (1) Creating an **orthography** for an unwritten contact language spoken by a multilingual community speaking different other languages with different orthographic systems;
- (2) Between the needs of the speech vs. the scientific community: the problem of addressing **different readership target groups** with different aims;
- (3) The centre and the periphery: The question of **exhaustivity** in terms of features as well as language variation along the creole continuum; the question of quantitative statements
- (4) Questions regarding the **empirical base**: The (pre-)selection of data with special attention to – sometimes contradicting – data from earlier fieldwork research.

References

Götze, Angelika / Lindenfelser, Siegwalt / Lipfert, Salome / Neumeier, Katharina / König, Werner / Maitz, Péter (2017): Documenting Unserdeutsch (Rabaul Creole German): A workshop report. In: Maitz, Péter / Volker, Craig A. (eds.): *Language Contact in the German Colonies: Papua New Guinea and beyond* (= Special issue of *Language and Linguistics in Melanesia. Journal of the Linguistic Society of Papua New Guinea*), 65–90.

Lindenfelser, Siegwalt / Maitz, Péter (2017): The creoleness of Unserdeutsch (Rabaul Creole German): A typological perspective. In: Maitz, Péter / Volker, Craig A. (eds.): *Language Contact in the German Colonies: Papua New Guinea and beyond* (= Special issue of *Language and Linguistics in Melanesia. Journal of the Linguistic Society of Papua New Guinea*), 91–142.

Maitz, Péter / Volker, Craig A. (2017): Documenting Unserdeutsch. Reversing colonial amnesia. In: *Journal of Pidgin and Creole Languages* 32.2 (Amsterdam & Philadelphia), 365–397.

Alexander Mankov

Department of Swedish, Gothenburg University;
 Royal Society for Swedish Culture Abroad (Sweden);
 St. Tikhon's University for the Humanities (Russia)

Riksföreningen Sverigekontakt, Dicksonsgatan 6, 412 56 Gothenburg, Sweden
 [mankov2017@gmail.com]

Towards a Description of the Dialect of Gammalsvenskby

The dialect of Gammalsvenskby is the only surviving Scandinavian dialect in the territory of the former Soviet Union. The village of Gammalsvenskby (its current Ukrainian name is Zmiyivka) is located in Cherson region, Ukraine. It was founded in 1782 by migrants from the island of Dagö (Estonian *Hiiumaa*), which then belonged to the Russian Empire. Historically, the dialect goes back to the Swedish dialects of Estonia. They developed as a result of East-Scandinavian expansion towards the East since 13–14th centuries. In contrast to Iceland and the Faroe Islands, which were colonised due to the Scandinavian expansion to the West a few centuries earlier, in the islands of the East Baltic Sea no uniform Scandinavian language has developed. All Swedish dialects of Estonia except for the dialect of Gammalsvenskby are now extinct. The number of the present-day speakers does not exceed 10 persons; the initial number of the migrants from Dagö was ca. 1000. The present-day speakers are not uniform in their knowledge of the dialect; there is a group of fluent speakers and semi-speakers. The former had the dialect as the main language in childhood and young age, the latter learnt it as adults while communicating orally with the fluent speakers. The main language in the village is now Surzhik (Russian-Ukrainian transition variety), all dialect speakers have a good practical knowledge of German and Standard Swedish.

From 2004 to 2013, I made ten trips to the village with the aim of documenting the dialect. Before this, the current state of the dialect was unexplored. I have developed a spelling system for the dialect and collected material sufficient for a detailed grammar and comprehensive dictionary. In April 2018, the Royal Society for Swedish Culture Abroad (Riksföreningen Sverigekontakt) in conjunction with Gothenburg University, Sweden, launched a project on making a description of the dialect. In my paper, I will study main features of phonetics and morphology of the dialect and discuss the principles of its description.

As for the phonetics, the dialect — compared with Standard Swedish — is characterised by the following features: non-aspirated *p, t, k*; frequent occurrence of [ɾ], retroflex-alveolar flap, which is an allophone of /l/; absence of the postalveolar *l*; dental *n < m* within morphemes (e.g. *bōn* ‘child’; cf. Sw. *barn*); palatalised [sʲ] < *hj* (*sjūn* ‘person’, cf. Sw. *hjon*); the combination [x:l] < *ssl* (*näxxlar* ‘nettle’); preservation of *g, k* in front of the front vowels; the absence of both short and long /y/, /ö/; the vowel [ɛ] corresponding to Sw. short *u*; presence of diphthongs; lack of word intonation; presence of not only long but also short stressed syllables.

As for the grammar, the dialect preserves the three-gender system of nouns and adjectives; features considerable variety of sandhi regularities in the definite singular form of nouns (particularly in the masculine); the inflectional classification of masculine and feminine nouns requires to take into account not only the plural but also the definite singular form; the plural form of adjectives depends on its syntactic position (attributive vs. non-attributive form); the perfect employs the verbs “have” and “be”.

Chasing the clause: the challenges of describing person indexing in Äiwoo

Åshild Næss, University of Oslo

The Oceanic language Äiwoo indexes the actor and, in some cases, the undergoer argument by affixes on the verb: prefixes for intransitive and actor-voice transitive verbs (1) and suffixes for undergoer-voice transitive verbs (2).

- (1) a. *I-ku-wä.*
1MIN-IPFV-go
'I'm going.'
- b. *I-ki-vängä* *sii.*
1MIN-IPFV-eat.A fish
'I'm eating fish.'
- (2) *I-togulo-nee-mu.*
PFV-hit.O-1MIN-2MIN
'I hit you.'

For the description of person indexing, Haspelmath (2013) proposes a distinction between gram-indexes, which obligatorily cooccur with a coreferent nominal (a conominal) in the same clause; cross-indexes, which allow but do not require a conominal; and pro-indexes, which do not permit a conominal.

These distinctions are intended to allow description of the linguistic facts without necessarily committing to a side in the discussion of the status of bound pronouns as either 'agreement' or 'argument' (e.g. Jelinek 1984, Bresnan and Mchombo 1987, Baker 1996, among others).

However, they hinge in turn on a no less problematic concept, namely that of a clause: How does one establish whether a nominal coreferent with the person index occurs within the same clause or not?

Most definitions of the concept 'clause' in the literature are fairly summary, e.g. "a syntactic unit consisting of a subject and a predicate" (Freidin 2012: 266) or "syntactic units packed under a single intonation contour" (Givón 2001: 355). Such definitions do not readily translate into tests for clause boundaries that can be applied to languages for which the status of person marking vs argument NPs is unresolved.

In this paper, I will discuss the problems arising in trying to establish clause boundaries in Äiwoo. Specifically, I will address the highly frequent deictic particles *lä* and =*Cä* which function to 'bracket' syntactic units, as in (3) (deictics in bold):

- (3) *Dä nyidäbu=dä, lä i-wo-lä=to=wä ngä numa lägä*
some day=some DIST PFV-go-out=now=DIST LOC reef empty
lä ki-tei=to=wä.
DIST IPFV-fish=now=DIST
'One day he went out at low tide and fished.'

In a head-marking language, how does one establish whether the entity bracketed by the deictics is a phrase or a clause? In other words, is the noun *sime* 'person, people' outside the clause in (4) by virtue of being to the left of the bracket, or does the clause include constituents outside the bracket? Consequently, is *lu-* 'they' a pro-index or a cross-index?

- (4) *Sime lä lu-po-lä-du-kä nyiva=kä.*
person DIST 3AUG-go-out-all-DIR:3 outside=DIST
'All the people went outside.'

I will address the function of the deictic particles in the structuring of Äiwoo discourse and the distribution of entities such as the noun *sime* ‘person’, the modifying element *du* ‘all’, and the coreferential index *lu-* ‘3rd person augmented’ across the constituents of the clause, and discuss to what extent ‘clause’ is actually a helpful analytical concept in understanding syntactic organisation in Äiwoo.

References

- Baker, Mark C. 1996. *The polysynthesis parameter*. New York: Oxford University Press.
- Bresnan, Joan, and Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chichewa. *Language* 63:4, 741-782.
- Freidin, Robert. 2012. *Syntax: Basic concepts and application*. Cambridge: Cambridge University Press.
- Givón, Talmy. 2001. *Syntax*. Volume II. Amsterdam: John Benjamins.
- Haspelmath, Martin. 2013. Argument indexing: a conceptual framework for the syntactic status of bound person forms. In Dik Bakker and Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 197-226. Hawthorne, NY: Walter de Gruyter.
- Jelinek, Eloise. 1984. Empty categories and non-configurational languages. *Natural Language and Linguistic Theory* 2, 39-76.

Abbreviations

A actor-voice verb, AUG augmented number, DIR directional, DIST distal, IPFV imperfective, LOC locative preposition, MIN minimal number, O undergoer-voice verb, PFV perfective,

Elatives of Kina Rutul

Polina Nasledskova, Michael Daniel

Linguistic Convergence Laboratory, HSE University, Moscow

Rutul is a Lezgian (East Caucasian) language spoken in southern Dagestan, Russia. As most of its sister languages, it has a wide system of spatial forms comprising two categories, localization (In ‘inside’, Apud ‘near’, Sub ‘under’, and some others) and orientation (static location, motion to, motion from, and some others). Wrt. to the category of orientation, East Caucasian spatial systems may be classified into tripartite, distinguishing between essive (static location), lative (motion towards) and elative (motion from), and bipartite, where essive/lative are merged in one form (usually unmarked) and opposed to the marked elative form. Rutul of Kina is an example of a bipartite system, with most elative forms containing a dedicated marker *-la* and the lative identical to the essive (unmarked). However, in segmental terms, essive, lative and elative are identical for In ‘inside’ and Apud ‘near’ localizations - (*-a* and *-da*) respectively.

The situation where spatial forms do not distinguish elative from lative and essive, present in some African languages (e. g. in Wan (Mande), see (Nikitina 2009)), seems to be very unusual for Dagestan. Sources suggest that some of the dialects of Rutul distinguish between In-Essive(Lative) and In-Elative by vowel quantity (Maxmudova 2002: 47; Ibragimov 2004). The language is unwritten, and not all Kina speakers are confident about the nature (or even presence) of the distinction. Perceptually, the difference was not always clear (except in very articulated style of pronunciation); it is also not clear whether the duration or intensity played a role in the distinction (if any).

To check this for Kina, we conveyed the following experiment. We elicited sentences with essive, allative and elative forms. Two types of contexts were used:

1. Noun in a spatial form in a medial position in the sentence
2. Noun in a spatial form in the end of the answer to a ‘Where?’ question.

The target contexts were supplemented by fillers, different speakers were given the same stimuli in different order, with contexts with phonetically similar forms separated from each other by other contexts and fillers. We then measured the absolute vowel length and intensity of the vowels and compared them within each speaker (see appendix for all measurements; some of the realizations had to be rejected on different grounds). The absolute vowel length was measured from the first to the last visible period. For intensity, the maximum value was taken. An overall number of recorded target contexts is 50. On average, /a/ in elatives appeared to be about 1.5 times longer than in essive and allative form. As expected, there was no consistent or significant difference between the essive and lative forms. Intensity does not seem to play any role, Pearson correlation between intensity and the type of the locative case is 0.1284, which is considered small. However, further statistical analysis is needed to support the conclusions.

In addition to these calculations, we are planning to carry out a perceptive experiment in the next field season (summer 2019). In this experiment, the subjects will be

given audio of the wordforms extracted from the stimuli described above and asked to judge what is, for these forms, the more appropriate spatial context.

Notably, there is some support that the weakening of the salience of the distinction between elative and essive/lative is structurally compensated by an emerging tendency to additionally use postpositional adverbs that disambiguate the forms by formally distinguishing essive/lative from elative.

References

- Nikitina, T. 2009. Subcategorization pattern and lexical meaning of motion verbs: A study of the source/goal ambiguity. *Linguistics*. 47.
- Ибрагимов Г. Х. 2004. Рутульский язык: Синхрония и диахрония. Махачкала: Изд. Дом «Народы Дагестана».
- Махмудова С. М. 2002. Грамматические классы слов и грамматические категории рутульского языка. (Текст): автореф. дис. на соиск. учен. степ. д-ра фил. наук (10.02.02) / Махмудова Светлана Мусаевна; ДГУ. Махачкала.

Appendix. Relative duration of essive, elative and lative across speakers

		Lative to essive	Elative to essive	Elative to lative
speaker1	'house'	0.8	NA	NA
speaker1	'forest'	NA	NA	NA
speaker1	'neighbour'	1.09	1.55	1.43
speaker 1	mean	0.95	NA	NA
speaker2	'house'	0.66	1.13	1.7
speaker2	'forest'	NA	NA	0.5
speaker2	'neighbour'	0.67	1.13	1.68
speaker2	mean	0,67	1.13	1.29
speaker3	'house'	NA	NA	NA
speaker3	'forest'	NA	NA	1.61
speaker3	'neighbour'	0.89	1.22	1.37
speaker3	mean	NA	NA	1,49
speaker4	'house'	NA	NA	1.26
speaker4	'forest'	NA	NA	2.05
speaker4	'neighbour'	1.57	2.24	1.42
speaker4	mean	NA	NA	1,58
speaker5	'house'	6.5	12.17	1.87
speaker5	'forest'	NA	NA	NA
speaker5	'neighbour'	NA	2.87	NA
speaker5	mean	NA	7,52	NA

speaker6	'house'	0.75	1.18	1.56
speaker6	'forest'	NA	NA	1.06
speaker6	'neighbour'	2.1	3.49	1.65
speaker6	mean	1,43	2,34	1,42
median	all speakers	0.89	1.55	1.56

The role of language variation in a grammar of threatened language without fixed standard: The case of Võro grammar

Helen Plado (University of Helsinki, Võro Institute)

Võro is Finnic language that is spoken in south-east Estonia. According to last census (2011), there are about 75000 Võro speakers. However, nowadays, there are no monolingual Võro speakers, and Võro language is under the very strong influence of Standard Estonian. Although books are published in Võro, it is used in printed media as well as it is taught in some schools in the area, its standard is still developing. Some general agreements about Võro standard are made; however, the wide range of variation in the language is accepted (for example also in printed media).

Võro descriptive grammar has been written in Võro Institute since 2017. As previous grammars of Võro (Wiedemann 2002 [1864] and Keem 1997) are rather short overviews, the new grammar will be addressed to wide readership that include native and new speakers on the one hand and language learners and teachers on the other hand, as well as linguists. Language community has been put into the foreground, hence the grammar will be written in Estonian.

In the presentation, I will argue that due to the lack of general standard language and wide readership, it is crucial to describe the variation of the language. Two types of variation will be discussed: areal/dialectal variation and generational/diachronic variation.

I claim, that the presenting of the areal variation in grammar of the threatened language has advantages for all the reader groups. However, it also raises the questions, e.g. is it possible to describe all the dialectal variation in the grammar, or should the grammar describe only the variation of the most essential features (e.g. well-known division of the marker of inessive case in Võro, where *-n* and *-h* are used in western and eastern dialects respectively)? To what extent grammar should describe highly characteristic features, that are spread in a quite small area? Which role corpora play in making the decisions about inclusion of the variation?

In the case of describing generational variation, the main issue in writing Võro descriptive grammar has been the inclusion of the nowadays spoken language, that is strongly influenced by the majority language. How to deal with the constructions following Estonian patterns, or being clearly influenced by Estonian? Should these be in the same row with the other variants or should these be presented differently, e.g. in remarks? In connection to the issue, the following question rises – is it reasonable or even possible to avoid the prescriptive approach completely in writing a descriptive grammar of a threatened language for the wide readership?

References

- Keem, Hella 1997. Võro keel. Tallinn: Emakeele Selts.
 Wiedemann, Ferdinand Johann 2002 [1864]. Versuch ueber den Werroehstnischen Dialekt. Tartu: Tartu Ülikool.

Olga Potanina
Nazarbayev University

Andrey Filchenko
Nazarbayev University

A Theory and Typology of Possession in Ob-Yenisei Languages

The project studies, from the typological and areal perspectives, a set of North-Western Siberian indigenous languages comprising the so called Ob-Yenisei linguistic area: Teleut, Tom Tatar, Chulym Turkic, Selkup, Khanty, Mansi, Ket, Nganasan, Enets, Nenets. These languages are all lesser described and highly endangered idioms, some numbering less than 5 speakers.

The project's concept stems from three facts: 1) there are no comprehensive descriptions of possession in the idioms of the project, but there are unpublished data on these highly endangered and lesser studied dialects, largely unavailable to academic community; 2) there are no exhaustive typological parameters defined for the cross-linguistic study of possession; 3) there is no theoretically and methodologically consistent typological description of the possession in the Ob-Yenisei languages taking into consideration areal factors and language contact.

Studies of possession in these languages have not been exhaustive, the existing works on the topic do not cover the wide range of functions and types of possessive constructions attested in these systems, while very few of these studies are performed within modern typological perspective. By focusing on creating a "typological portrait" of possession in Ob-Yenisei languages, the project performs description, within a uniform theoretical and methodological perspective, of a wide range of possessive constructions, in structurally diverse, genetically related / remotely related / unrelated languages of the area. Thus, the project on one hand, produces a descriptive study of Ob-Yenisei languages, informed by the modern typological theory, and taking into consideration most recent typological parameters relevant for the study of possession. On the other hand, the project pursues to design a typological questionnaire for the study of possession (akin to MPI Leipzig questionnaires). Objectives of the project include producing maximally exhaustive typology of both adnominal and predicative possession, revealing complexity of the conceptual and grammatical domain of possession. Thus, the project is concerned not only with mere illustration a wide range of morphosyntactic strategies employed by these languages, but also with describing and interpreting the full diversity of possessive constructions in terms of interrelation of various parameters within general typological perspective, grammaticalization theory, information structure theory, utilizing typological parameters elaborated for description of adnominal and predicative possession by M. Koptjevskaja-Tamm L.Stassen, B.Heine, among others. Being a universal category, possession does not have universal ways of manifesting possessive relations in languages (various lexical, morphological, syntactic means for marking possessive relations). The notion of possession remains ambiguous. Besides formal variation, possessive constructions have various functions and meanings: ownership, part-whole, kinship, associative possession, and pragmatic functions such as definiteness/indefiniteness, specificity, salience.

The project is at its initial stage, and seeks to review in uniformly both, intra- and extra-linguistic principles that condition the whole variety of the possessive constructions attested in the languages of the area, establishing the parameters that would be valid for cross-linguistic comparison. Upon analysing a restricted set of language data (6 languages), the following **tendencies** may be observed: (i) Possessor (PR) and Possessee (PE) typically comprise single noun phrase; (ii) Possessor tends to be animate/human; (iii) Possessor is coded by a personal pronoun or a proper noun; (iv) Possessor has topicality status; (v) pronominal Possessor is often expressed differently from the Possessor of other grammatical classes; (vi) type of possessive construction depends on whether the Possessor is human/animate, highly-referential/non-referential, alienability of possessive relation; (vii) the notion of 'ownership' is subject to local cross-cultural variation, specifying variously the sort of entities that can be 'owned'; (viii) cultural specific features define the domain of inalienable possession; (ix) for the predicate possession, formal

parameters valid for cross-linguistic comparison include: a) treatment of nominal and pronominal arguments *vis-a-vis* subjects and objects (identical or differential); b) dominant *Habeo* vs. *Esse* pattern in the language; c) formal coding of arguments; d) use of “possession” verb with inalienably possessed items (possible only for alienable?); e) sensitivity to definiteness and focality of Possessee and Possessor, f) sensitivity to temporality of possession. Possessive markers also function as cross-reference markers in narratives, used for discourse coherence, participant identification and disambiguation. It is also apparent that the negation strategies may vary conditioned by pragmatic status of the referents (for example, Locational possessive constructions may follow existential negation pattern).

Examples:

[I] *Adnominal possession* (based on (Croft; Koptjevskaja-Tamm; Aikhenvald)):

1.1.) Vasyugan Khanty (Potanina, Filchenko):

(a) *mä ap-am* (b) *tuy-i* *ʃi* *apajy-ən* *ʃaj*
 1SG father-1SG/SG carry-IMPR.2SG DET grandfather-2SG/SG stuff
 My father. (inalienable) Carry that (your) grandfather’s stuff. (alienable ‘stuff’)

(c) *wajaj lök(-əl)* *män-ä* *pämiltäl-tə*
 animal footprint-3SG/SG 1SG-ILL show-PST0.3SG/SG
 (S)he showed me animal foot-print. (optional PX - alienable (?)) - footprint made by animal

(d) *mən-nə kür-t-əm* *loŋə-ta* *Zoltan* *nipik*
 1SG-LOC cannot-PRS-1SG read-INF Zoltan book
 I cannot read Zoltan’s book. (book written by Zoltan – as opposed to owned by Zoltan)

(e) *emtər qanəŋ* (f) *qat jor* (g) *potinka pelək*
 lake middle house middle “shoe” half
 Middle of the lake. House floor. One shoe (half a pair). (associative/relational)

(h) *aj qu küŋkül-tə tayi-l-oŋ* *joy-pa* *ju-wəl*
 little man hunt-NPP place-3SG-ABL home-ILL come-PRS.3SG
 The young man came home from hunting. (place of regular hunts – nominalization)

(i) *pəy-l jayənt-əkətə-yən* *ajri juy wej morəmta-yən*
 son-3SG row-INCH-PST0.3SG canoe tree handle break-PST0.3SG
 His son started rowing when **his oar** broke. (possessive relation recoverable from context /topical)

1.2.) Southern/Central Selkup (Kim, Kovylin, Budzisch):

a) *üčedie-li-ka-n* *amba-d* *ur-u-k* *čur-e-li-mba*
 child-DIM-DIM-GEN mother-3SG force-EP-ADV cry-EP-INCH-PST.NAR.3SG
 The child’s mother began to cry strongly. (GEN-marked possessor - inalienable possession)

b) *qana-t qob*
 dog-GEN hide
 Dog hide. (GEN-marked PR - alienable possession - hide owned by a human as object)

1.3) Ket (Krjukova, Nefedov):

a) *aqta ke²d-da* *battad* b) *bū-da* *sóòm*
 good man-3SG.PX.M forehead he-3SG.PX.M arrow
 Forehead of a good man. (nominal PR) His arrow (pronominal PR)

- c) *da-ām-diŋa* *uska d-o-n-a-dij*
3SG.PX.M-mother-F.DAT home 3M.SBJ-PST-PST-3SBJ.COREF-come
 To (own) mother's place he came. (PR - not coded explicitly).
- d) *n-asl-n* e) *qam-t* *kub-'diŋa*
 PX.PL.ANIM-ski-PL arrow-PX.INAN end-DAT
 Their skis. (head-marking) To the end of the arrow. (dependent-marking)
- f) *destul* [des-d-ul] g) *desul* [des-ul]
 eye-LINK-water eye-water
 Tear. (linking element-PX.3SG) Tear. (no linking element)

[II] Predicate Possession (based on (Heine, Stassen):

2.1) Vasyugan Khanty (Gulja, Honti, Potanina, Filchenko)

- a) *n'oyos moq-qən taja-wəlt* b) *mä läwət al taja-l-əm*
 sable baby-DU have-PRS.3PL 1SG seven year have-PRS-1SG
 They keep two little baby-sables. I am 7 years old. (*habeo* constructions)
- c) *qaq-əl əj-ni toja-ɣən* d) *mä nipik-əl'i taja-ɣal-əm*
 yo.brother-3SG little-woman have-PST0.3SG 1SG book-DIM have-PST1-1SG
 He had a daughter. I had a good book. (*habeo* constructions)
- e) *mən-nə əj nəws-am patron wəl-kal*
 1SG-LOC one misfire-PP cartridge be-PST.3SG
 I had a cartridge that misfired. (Locational possessive construction: PE - subject; PR - LOC-marked)
- f) (*jəɣ-ən*) *öyi-t əntim wəl-qal* (?) g) *jüy öyi-t əntə taja-qal*
 3PL-LOC daughter-PL NEGex be-PST.3SG 3SG daughter-PL NEG keep-PST.3SG
 They had no daughters. (preferred NEG construction) (S)he did not have/keep a daughter.
- h) *mən-nə nipik wəl-wəl* i) *nipik mən-nə wəl-wəl*
 1SG-LOC book be-PRS.3SG book 1SG-LOC be-PRS.1SG
 I have a book. (PR-Topic / Topic-possessive) I have a book. (PE-Topic / Locational possessive)

2.2.) Southern/Central Selkup (Kim, Kovylin, Budzisch):

- a) *n'ab-i-nan naj šiði haj e-ja.*
 duck-EP-LOC EMPH two eye be-PRS.3SG
 The duck has two eyes. (Locational possessive construction: PE - subject; PR - LOC-marked)
- b) *mat kibajče-m komde-l čö-ye (ek).*
 1SG.GEN boy-1SG gold-ADJ belt-COM **be-3SG.sub**
 My fiancé has a golden belt. (Comitative possessive construction: PR-subject; PE- COM-marked)
- c) *nadak neb warə-nd-a.*
 girl duck keep-EVID.PRS-3SG.sub
 The girl keeps a duck. (*Habeo* construction: PR – subject/topic; PE – unmarked/focus).

2.3) Ket (Krejnovich, Krjukova, Nefedov):

- a) *abanja du'd usaj* b) *da ām-diŋta qāk tààn*
1SG.ADSS awl **be** 3SG.M.PX mother-F.ADSS five finger.PL
 I have an awl. ((non)past = *Esse* pattern) His mother has five fingers. (no *Esse* verb)
 ((2.3. a) - explicit *Esse* - Predicate-Focus vs. non-explicit *Esse* - PE-Focus (2.3.b))

- c) *Huʷn-diŋta* *kəʔd* *obilda-n*
 Khun-F.ADSS children “was”-PL
 Khun had children. (*obilda* Russian loan (cf. ‘был’ Georg, 2007: 314))
- d) *ēn* *ād* *tib-di-bed* e) *don-u-k-d/a-qon*
 now 1SG dog-1SG.SBJ-own knife-1SG.SBJ-become.PST
 Now I have a dog. ((non)past = *Habeo* pattern) I have got a knife. ((non)past = *become* pattern)

((2.3.d) - *own* construction and (2.3.e) - *become* construction cannot have transitive variant with PE - outside of the verbal predicate; both are PR-Topic constructions).

((2.3.c), (2.3.d) and (2.3.e) - are extremely rare - possible late innovation (Verner 1974: 44)).

[III] Possessive markers coding pragmatic relations:

3.1) Vasyugan Khanty (Potanina, Filchenko)

- a) *məŋ-ən* *qol-kas-əw* *tuyi* *tʃel'-m-əl*
 1PI-LOC hear-PST3-1PL thunderstorm cry-PP-3SG
 We heard the thunderstorm rattling. (Possessive markers - pragmatic relations)

3.2) Southern Selkup (Kim, Kovylin)

- a) *era-t* *surə-(j)-lje* *paljze-qu-ŋ.* b) *tjel-t* *qarət* *tjelə-ŋ.*
 oldman-PX.3SG animal-VBLz-CVB go-HAB-3SG sun-PX.3SG morning rise-3SG
 The oldman goes hunting. (PX.3SG Topic-marker) The sun rises in the morning (PX.3SG-unique)
- c) *ewe-l* *tabi-m* *ogolčoldzi-mb-an* *olo-n-d* *Ioanni-m* *Lostikudim* *madir-gu*
 mother-2SG 3SG-ACC teach-PSTN-3SG head-ACC-3SG John-ACC Baptist ask-INF
 (Her) mother taught her to ask for the head of John the Baptist. (PX.2SG – Topic-marker)

3.3) Northern Samoyedic (Siegl, Gusev, Wagner-Nagy)

- a) *mana-jet* *onai* *ne-r* *mana*
 say.3SG-EMPH real woman-2SG say.3SG
 So she said, this Enets women said. (Forest Enets - PX.2SG Topic-marker)
- b) *təti* *Muŋkat'i-rə* *kobtuatu* *təiśüə,* *ńerbiaʔku.*
 that Munkachi-2SG girl.3SG have.PST.3SG live.PST.3SG
 This Enets had a daughter, a little girl. (Nganasan - PX.2SG Topic-marker)
- c) *čiki mä-l* *nenado* *nerha* d) *čiki-r* *meju* *mäʔ*
 this house-2SG visible stand.3SG this-2SG new house
 This house seems to stand in plain view. This is a new house.
 (Tundra Nenets - PX.2SG Topic-marker) (Forest Enets - PX.2SG Topic-marker)

References:

- Большой словарь кетского языка. Под ред. Е. Г. Которовой, А. В. Нефедова. LINCOM. 2015. Т.1. 430 p.
- Вернер Г. К. Реликтовые признаки активного строя в кетском языке // Вопросы языкознания. 1974. №1. pp.34-45.
- Ким-Малони А. А., Ковылин С. В. Таксономия языковых средств выражения посессивности в селькупском языке. //Томский журнал лингвистических и антропологических исследований. 2015. № 4(10). pp.36-44.

- Которова Е.Г., Нефедов А. В. Типологические характеристики кетского языка: вершинное или зависимостное маркирование? // Вопросы языкознания. 2006. №5. pp.43-56.
- Крюкова Е.А. Именные посессивные конструкции в кетском языке // Томский журнал лингвистических и антропологических исследований. 2015. № 4(10). pp.55-61.
- Крюкова Е.А. Предикативные посессивные конструкции в кетском языке (корпусное исследование) // Томский журнал лингвистических и антропологических исследований. 2016. № 4(14). pp.22-36.
- Нефёдов А. Отрицание неглагольных предикатов в кетском языке // Томский журнал лингвистических и антропологических исследований. 2018. №1(19). pp.20-40.
- Потанина О. С. О проекте «Типология категории посессивности на материале языков обско-енисейского ареала» //Томский журнал лингвистических и антропологических исследований. 2015. № 4(10). pp.18-23.
- Budzisch J. Possessive constructions in Southern Selkup dialects.// Tomsk Journal of Linguistics and Anthropology, 2015. №.4(10). pp.45-50.
- Georg S. A Descriptive Grammar of Ket (Yenisei-Ostyak) GLOBAL ORIENTAL LTD, 2007. 328p.
- Croft W. Typology and Universals (2nd ed.). Cambridge, Cambridge University Press. 2003.
- Filchenko A. Possessive Constructions in Eastern Khanty. Tomskiy zhurnal lingvisticheskikh i antropologicheskikh issledovaniy – Tomsk Journal of Linguistics and Anthropology, 2014. №.1(3). pp.83-95.
- Heine B. Possession: Cognitive Sources, Forces, and Grammaticalization. Cambridge University Press. 1997. 292p.
- Honti L. 'HABERE' po-ural'sky. Linguistica Uralica XLIV. 2008, №3, pp.161-177.
- Koptjevskaja-Tamm M. Possessive Noun Phrases in the Languages of Europe. In F. Planck (Ed.), Noun Phrase Structure in the Languages of Europe. Berlin, Mouton de Gruyter. 2003. pp.621-722.
- Nikolaeva I. Possessive affixes as markers of information structuring: Evidence from Uralic. Comrie, Bernard, Valery Solovyev, and Pirkko Suihkonen (eds.). 2003.
- Siegl F. The Structure of Noun Phrases with Referential PX.2P in Northern Samoyedic. //Tomsk Journal of Linguistics and Anthropology. 2015. №.1(7). pp.21-31.
- Stassen L. Predicative Possession. Oxford: Oxford University Press. 2009.
- Wagner-Nagy B. Possessive Constructions in Nganasan. Tomskiy zhurnal lingvisticheskikh i antropologicheskikh issledovaniy – Tomsk Journal of Linguistics and Anthropology. 2014. №.1(3). pp.76-82.

Language descriptions for teachers of minority languages

Maria Pupynina (Institute for Linguistic Studies, St. Petersburg)

The talk is dedicated to the type of grammar descriptions that could be used to train minority language teachers and satisfy the needs of native speakers interested in language structure. Mostly, the situation in Russian Federation will be considered.

The majority of grammars of minority languages of Russia have reference grammars published in the Soviet times. Today, in the end of the second decade of the 21st century, these descriptions are still used as teaching materials (though they were not written as such) in universities where teachers of minority languages are trained. Most of these descriptions, as reported by many modern linguists, are heavy heritage of Russian-based linguistic paradigm, old-fashioned in terminology and over-complicated. Besides, all Soviet time grammars are definitely form-oriented.

For example, a 2-volume Chukchi grammar was published in 1961 and 1977 by soviet scholar P.Y. Skorik (non-native speaker). It is a detailed description based mainly on elicitation material. It abounds in language examples and contains information on many minor, rarely found affixes. This work is used as an important source material by modern language specialists, but it is too overcomplicated for future teachers. During our continuous talks with the students of language department of Institute of the Peoples of the North (St. Petersburg), they admitted that this grammar description is absolutely non-understandable for them.

For some of minority languages of Russia there are also newer, typologically-oriented grammars (this can be said about Chukchi, Alutor, Yukaghir, Nenets, Votic and other languages). Unlike pioneer grammars which were aimed to describe the language per se, typologically oriented grammars facilitate the comparison between the described language and other languages. In many cases, their authors try to maintain a nice balance between form and function. It is true for Chukchi grammar written by Michael Dunn in English (1999), which is a highly professional, well-written description based mainly on text material. However, this type of grammars also do not fit minority language teachers' needs. Such descriptions contain not easily explainable concepts and terminology which are not taught in universities where these teachers are trained.

We suppose that a good way to describe the grammar for minority language teachers is through the comparison of linguistic features (especially dissimilar ones) between the described language and the dominant idiom. For Russia, the comparison would embrace minority native language and Russian. It seems that emphasizing linguistic differences will help to explain (if needed) difficult concepts such as 'transitive paradigm', 'ergativity', 'singulative', 'person hierarchy' etc. In such description it would also be easier to clarify general characteristics of language system. For Chukchi, for example, comparing Russian (analytic) way of expressing desiderative meaning to Chukchi desiderative circumfix could be used to underline a more synthetic nature of Chukchi language compared to Russian, etc.

Creating such kind of description would help to make grammar more functionally-oriented and, at the same time, would facilitate non-professional's understanding of language structure. One could object that comparison with the emphasis on dissimilar features is not enough to describe the language 'entirely'. However, explanation of at least several concepts is good for the beginning, and still, the Soviet and typologically-oriented grammars are accessible as the next step of theoretical training.

Grammar Writing and Diachronic Typology: Friends or Foes?
 Jorge Emilio Rosés Labrada
 University of Alberta

Grammar writing has recently seen an ever-increasing number of works dedicated to its theorization (see Payne & Weber 2006; Ameka, Dench & Evans 2006; Nakayama & Rice 2014, *inter alia*). However, the role of diachrony in grammar writing has remained relatively undertheorized (but see Rankin 2006 and Post 2013), being even often completely left out of work discussing the methodology of grammar writing (e.g. Nikolaeva 2016). As Rankin (2006:527) mentions, it is likely that this stems from the “Saussurian prohibition against mixing synchrony and diachrony”. By and large, typology has also focused on the synchronic distribution of particular features or patterns as well as on synchronic explanations for observed distributions but recent work (e.g. Cristofaro 2013, 2017; Hendery 2012; *inter alia*) has shown the importance of considering the diachrony of linguistic properties. In this presentation, I argue that it is crucial that grammar writers investigate, and include in their grammars, the evolution of features and constructions of the language under study by focusing on differential object marking (DOM) in Mako (Sáliban, Venezuela) as a case study.

In Mako, the P argument of a transitive verb and the T and R arguments of a ditransitive verb being optionally marked with a suffix *-ni* (Rosés Labrada 2015:344-7). Puzzlingly, this same suffix is used to obligatorily mark verbal arguments with the semantic roles of instrument, goal and location (*ibid*, 347-51). This leads Rosés Labrada to talk about a single non-subject suffix *-ni*. The diachronic typology of DOM would suggest that Mako DOM may have followed a similar pattern to Spanish DOM. However, comparison with Piaroa and Sáliba, Mako’s closest relatives as well as with more distantly-related Jodí suggests that perhaps a better analysis would be to posit two homophonous suffixes. Both Piaroa and Sáliba exhibit a similar pattern of DOM to that of Mako. The relevant markers are *-ri/-ru* in Piaroa (Krute 1989) and *-ri* in Sáliba (Morse & Frank 1997), both cognate with the Mako marker (see Rosés Labrada, accepted). On the other hand, arguments with the semantic roles of instrument and location in Piaroa and Sáliba are marked with the suffixes *-næ* and *-na*, which are not cognate with the Mako marker since *i:æ:a* is not a regular sound correspondence. Adding Jodí data to the comparison serves to illuminate this puzzle. In Jodí, P and T can be marked with a *-ni* suffix if animate (Quatra 2008) and R arguments with *-likī* while arguments with an instrumental locative role can be marked with *-ni* and arguments with a locative semantic role can be marked with a suffix variably realized as *-ni*, *-nē*, or *-na*. This comparison suggests that the common ancestor of Jodí-Sáliban had two distinct markers: **-ri* for animate P, T and/or R and a **-nV* suffix for instrumentals and locatives.

This work thus argues that looking at the comparative data can result in a better analysis for the language under study and at the same time, advance (or at least not hinder) efforts to understand the evolution of DOM.

Selected references

- Ameka, F. K., Dench, A., Evans, N. (eds). 2006. *Catching Language. The Standing Challenge of Grammar Writing*. Trends in Linguistics 167. Berlin: de Gruyter Mouton.
- Cristofaro, S. 2013. The referential hierarchy: Reviewing the evidence in diachronic perspective. In Dik Bakker & Martin Haspelmath (eds.), *Languages Across Boundaries: Studies in Memory of Anna Siewierska*, 69–93. Berlin: dGM
- Nikolaeva, I. 2015. Reference Grammars. In Ungeheuer G. & Wiegand H. (eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK), Volume 3*, 2036-2063. Berlin: dGM.
- Rankin, R. 2006. The interplay of synchronic and diachronic discovery in Siouan grammar-writing. In Ameka, Dench, & Evans; pp. 527-548.
- Rosés Labrada, J.E. 2015. *The Mako Language: Vitality, Grammar and Classification*. PhD dissertation: The University of Western Ontario & Université Lumière-Lyon 2.

An overview of the progress and challenges of a Forest Nenets grammar project

Tapani Salminen (University of Helsinki / Kone Foundation)

This presentation discusses the prospects and problems of a recently established project aimed at describing Forest Nenets, a language of the Nenets subgroup of the Samoyed branch of the Uralic family spoken in north-western Siberia. In comparison with the Tundra Nenets language, from which it differs almost as greatly as Dutch from English, Forest Nenets is severely understudied and undercultivated as well as all but unrecognized. Specific issues related to the status and structure of Forest Nenets include the following: (i) The corpus of Forest Nenets materials, whether publications written by Forest Nenets authors or collections made by professional linguists, is extremely limited. What would be the means to remedy the lack of statistically significant databases, or what would be the most effective measures for combining corpus building and grammar writing? (ii) Since Tundra Nenets has been researched much more thoroughly than Forest Nenets, to what extent may the results of Tundra Nenets studies be adopted and profited from in the description of Forest Nenets, and what are the potential risks in doing so? Or more generally, what are the limits of comparative and contrastive approaches in a descriptive grammar? (iii) Besides dialect differences typical of any language, Forest Nenets is characterized by a high level of free variation, to the extent that a single speaker may employ not two but three distinct variants for some of the most frequent verb forms. Both practical and theoretical aspects of variation warrant further discussion on this point. (iv) Forest Nenets is not unique in possessing a rich and complex system of derivational morphology, but the question remains whether the description of derivation should be based on an annotated list of derivational processes, typical of many descriptive grammars, or whether derivational morphology might be better described in a manner that would resemble the description of inflection, i.e., involving rules and paradigms, even when derivation, unlike inflection, was characterized by opaqueness and lacunae. (v) One of the pledges of the Forest Nenets grammar project is that it intends to integrate grammar with texts and lexicon. This is, however, easier said than done, not only because of the practical problems mentioned above but also due to more fundamental issues concerned with the structure of a descriptive grammar. (vi) Finally, any descriptive grammar should presumably strive for a balanced description of phonology, morphology, and syntax. This is another goal not so easy to achieve, and some of the current limitations in the practices of descriptive linguistics and language typology will be discussed in the presentation.

Functions of participials and gerunds in Northern Khanty

According to Northern Khanty grammars, there are two participials, one expressing present time with the suffix *-ti* (being homonymous with the infinitive ending) and the other expressing past time with the suffix *-əm*. In Khanty, there are not many subordinate clauses or conjunctions and they have developed under Russian influence. In Northern Khanty, the old Uralian way to use shortened clauses or participial phrases formed by means of participial suffixes, possessive suffixes and locative case ending (*-ən*) is still productive. The participial forms expressing temporal simultaneous or not coincident event occur quite often in texts, e.g. *mǎn-t-em-ən* ‘when I was going’ (go-PRES.PTCL-PX1SG-LOC, *mǎn-ti* present participle) or *mǎn-m-em-ən* ‘after I was gone’ (go-PAST.PTCL-PX1SG-LOC, *mǎn-əm* past participle). Another alternative is to use postpositions such as *artən* ‘when, at the time’, *sat* or *sati* ‘after’, e.g. *noχər jetšǎ-t-al artən* ‘when the nuts (sg, of the Siberian pine) ripen’ (Solovar 2006: 27), *mǎn-t-em sat* ‘when I’m going’, *mǎn-t-al sati il pit-əs* ‘when (s)he was going (s)he fell’ (Solovar 2006: 245).

There is also a gerund, a verbal adverb, formed with full productivity from verbs by means of the suffix *-man*. Usually the gerund expresses simultaneous doing, e.g. *loj-man werat-l* ‘(s)he works standing’ (stand-GER work-PRES3SG) or *oməs-man werat-l* ‘(s)he works sitting’ (sit-GER work-PRES3SG) (Onyina 2009:47). It is not difficult to find in the same source examples of two dissimilar uses. On the one hand an example can be well suited for the description: *ewi ari-man šoš-əs* ‘the girl walked singing’ (Němysova 1988: 126), and on the other hand the next example can unexpectedly show different time levels: *iki molupsi lemat-man antup antup-man kamən tut+juχ sewr-əs* ‘after dressing the malitsa and fastening the belt the man chopped outside firewood’ (Němysova 1988: 126). These ranges of common use of participials and gerunds have hardly ever been studied. The knowledge of syntax and terminology of some other Finno-Ugrian language would greatly benefit further studies of Khanty syntax.

- Němysova, E. A. (ed.) 1988: *Hantyjskij jazyk. Učebnik dlja učaščihsja pedagogičeskih učilišč.* Leningrad: Prosveščenie.
- Onyina, Szofia 2009: *Szinjai hanti társalgási szótár (nyelvtani vázlattal és szójegyzékkal).* – Budapesti finnugor füzetek 20. Budapest.
- Solovar, V. N. 2006: *Hantyjsko-russkij slovar'.* Sankt-Peterburg: OOO Mirall.

Linguistic variation: a challenge for describing Kanashi phonology

Anju Saxena (Uppsala University)

Anna Sjöberg (Uppsala University)

Padam Sagar (Nako)

Lars Borin (University of Gothenburg)

Phonetics and phonology have traditionally received less attention in grammar writing than other aspects of language form. In a survey of twenty grammars published between 1989 and 2000 Maddieson (2002) found that only five of these devote more than one tenth of their number of pages to phonology or phonetics and that many do not include a phonetical description at all beyond giving phonemes their place in a table, which, as Maddieson points out, is not always enough for an understanding of the sound system of the language. In the same vein, Blevins (2007) points out that a better understanding of the sound system of a language may help us in uncovering or clarifying genetic relationships. She also emphasizes the importance of phonetic and phonological documentation for language revitalization.

While to some scholars (e.g. Ladefoged and Maddieson 1996) there is no essential difference between studying the phonetics of endangered languages and non-endangered languages, which is no doubt largely true, there are still some ways in which the sound systems of dying or endangered languages may differ from those of more ‘healthy’ languages. For instance, there may be variation in an endangered language without any apparent linguistic explanation. (Palosaari & Campbell 2011; Cook 1989). This may be the result of previously obligatory phonological processes being lost, giving rise to free variation. Alternatively, it could be due to several layers of language contact with the same set of dominant languages during the history of the language (Andersen 1982).

Based on our fieldwork data, in this presentation we will describe observed phonetic variation at the lexical level among speakers of Kanashi (xns), a “definitely endangered” (UNESCO) unwritten and almost undescribed Tibeto-Burman (TB) language. It has about 1,500 speakers (*Ethnologue*, Simons and Fennig 2018). Kanashi is spoken only in one village – Malana (Kullu district in the state of Himachal Pradesh in India), while in the surrounding villages (unrelated) Indo-Aryan (IA) languages are spoken.

We will relate this observed phonetic variation to challenges that one faces when analyzing the data with an aim to describe the phonology of Kanashi. Practical challenges concern issues such as classifying the lexicon into inherited and borrowed items, but also, in the case of a multi-layered contact situation with much variation between speakers, recognising which items (or forms) are used by which speakers depending on social variables (e.g. age or literacy). This sociolinguistic viewpoint may not be foremost in a researcher’s mind when doing preliminary phonetic and phonological fieldwork but as the situation in Kanashi shows, this may be necessary in order to answer basic questions regarding a language’s phonology, for example, deciding the phoneme inventory of Kanashi. Among the IA loanwords we find two layers: (i) those which apparently were borrowed long ago and (ii) more recent borrowings. The first set have been adapted to the TB core of Kanashi, for example, many borrowed IA nouns and adjectives in Kanashi end in *-aŋ* or *-iŋ*, where their IA counterparts do not have this ending. Occasionally both the adapted and non-adapted form of the same lexical item are found in our material – sometimes in the speech of the same speaker and at

other times, in data provided by different speakers. Phonology traditionally treats each language as a closed system with its own set of phonemes and phonological rules. In a language such as Kanashi with much borrowing, this boundary may be less clear.

References

- Andersen, R. W. 1982. "Determining the linguistic attributes of language attrition". In: R. Lambert & B. Freed (eds.), *The loss of language skills*. 83–118. Rowley, MA: Newbury House.
- Blevins, J. 2007, "Endangered Sound Patterns: Three Perspectives on Theory and Description", *Language Documentation & Conservation*, 1(1).
- Cook, E. 1989. "Is Phonology going haywire in dying languages? Phonological variations in Chipewyan and Sarce". *Language in Society*, **18**(2): 235–255.
- Ladefoged, P. & I. Maddieson. 1996. "Recording the phonetic structure of endangered languages."
- Maddieson, I. 2002. "Phonetics in the Field". *Berkeley Linguistics Society 28: General Session and Parasession on Field Linguistics*.
- Palosaari, N. & L. Campbell. 2011. "Structural aspect of language endangerment". In; P. Austin & J. Sallabank (eds.) *The Cambridge handbook of endangered languages*, Cambridge: Cambridge University Press.
- Simons, G. F. & C. D. Fennig (eds.). 2018. *Ethnologue: Languages of the world*, Twenty-first edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>

Documentation: When Inflection Meets Tone

Ronald P. Schaefer and Francis O. Egbokhare
Southern Illinois University Edwardsville and University of Ibadan

Minority Niger Congo languages in West Africa present unique challenges for typological investigation and language documentation. Chief among these is the feature of tone and its relation to categories of inflectional morphology. Compared to genetically related Bantu languages and their agglutinating systems (Hyman 2004, Good 2012), these languages are inflectionally impoverished, at least with regard to their segmental morphemes.

Recent analysis of the tone-morphosyntax interface (Heath and McPherson 2013, McPherson 2014) combined with earlier characterization of inflectional typology (Booji 1993) provides a fruitful direction for grammatical documentation. In the instance of Emai, an Edoid language of southern Nigeria (Elugbe 1989, Williamson and Blench 2000), inflection of the inherent type (tense, aspect) shows segment plus tone form, while the contextual type (agreement) appears as contrasting melodies on lexical subject.

In Emai, tonal character across the subject verb relation varies, leading to instances of agreement vs anti-agreement. The primary conditioning factor is mood and its variable expression of tense, i.e. indicative vs subjunctive. Serving as a host for tense is a subject clitic consisting of a vowel with variable tone. In the indicative, this clitic, whether overt or covert, controls tonal overlay. Distal tense-aspect classes (habitual, past perfect, prospective predictive) reveal a {High} overlay on subject lexical tone, while proximal classes (continuous, present perfect, prospective anticipative) remain inert to overlay, i.e. subject lexical tone is retained. In the subjunctive (negation, conative, imperative, hortative), category subtypes manifest low tone and the accompanying subject clitic exhibits uniform high. However, the subject clitic in the subjunctive, which is zero marked for tense, does not prompt overlay. It requires lack of tone agreement vis-à-vis lexical subject, i.e. anti-agreement.

The principle feature of the indicative is agreement between tonal melody of a lexical subject and a tense bearing subject clitic (SC), which can be overt or covert. For distal relations, the tonal melodies of SC and lexical subject exhibit a uniform right edge. When the clitic \acute{o} is /H/, it prompts a {H} overlay on the melody of a lexical subject like low tone \grave{o} *‘Oje,’* e.g. $\acute{o}\acute{j}\acute{e}\acute{o}\grave{o}\grave{e}\acute{e}\acute{m}\acute{a}$ [Oje SC.distal HAB eat yam] ‘Oje eats yam.’ When the clitic \grave{o} is /L/, there is no overlay on subject lexical tone, e.g. $\grave{o}\acute{j}\acute{e}\grave{o}\grave{o}\grave{e}\acute{e}\acute{m}\acute{a}$ [Oje SC.proximal CONT eat yam] ‘Oje is eating yam.’ Similarly, when a covert subject clitic $\langle\acute{o}\rangle$ is /H/, as in past perfect, it prompts a {H} overlay on the melody of a lexical subject, e.g. $\acute{o}\acute{j}\acute{e}\langle\acute{o}\rangle\grave{o}\acute{e}\acute{e}\acute{m}\acute{a}$ [Oje <SC.distal PAP> eat yam] ‘Oje ate yam.’ When a covert SC $\langle\grave{o}\rangle$ is /L/, as in present perfect, no overlay occurs on subject melody, e.g. $\grave{o}\acute{j}\acute{e}\langle\grave{o}\rangle\acute{o}\acute{e}\acute{e}\acute{m}\acute{a}$ [Oje <SC.proximal PRP> eat yam] ‘Oje has eaten yam.’

For anti-agreement in the subjunctive, its principle feature is a lack of agreement between the tonal melody of subject clitic and lexical subject. The clitic is uniformly expressed with a /H/ tone, while the lexical subject retains its lexical melody, which is right edge /L/. The lexical subject precedes a high SC and a low tone subjunctive category, e.g. $\acute{i}\grave{i}$ of negation, e.g. $\grave{o}\acute{j}\acute{e}\acute{i}\grave{i}\grave{e}\acute{e}\acute{m}\acute{a}$ [Oje SC NEG eat yam] ‘Oje did not eat yam’; conative, e.g. $\grave{o}\acute{j}\acute{e}\acute{o}\acute{o}\acute{e}\acute{e}\acute{m}\acute{a}$ [Oje SC

CON eat the yam] ‘Oje has gone to eat the yam.’ We conclude by emphasizing how syntagmatic aspects of tonal variation need to be assessed for possible correlations with inflectional categories, especially as this pertains to grammatical description of underdescribed languages.

Finding reader-friendly ways to incorporate syntax into a grammar of Tuparí

Adam Roth Singerman, University of Chicago

adamsingerman@uchicago.edu

How can grammatical descriptions aimed at a wide audience incorporate information on syntactic structure? LaPolla and Poa (2006), among others, advocate describing word order phenomena in linear terms; Rice (2006), in contrast, argues that analytic questions drawn from formal theory can enhance descriptions of understudied and endangered languages. This debate is made more complicated by the fact that reference grammars must satisfy various readerships (Mithun 2007; Noonan 2007; Rice 2007). No one analytic framework will be familiar to all readers, especially given persistent disagreement about how syntactic phenomena ought to be modeled (see, for example, the many different views in Borsley and Börjars 2011 and Heine and Narrog 2015).

This talk addresses the issue of syntax-in-reference-grammars by discussing Tuparí, an endangered Tupían language spoken by some 350 people in the Brazilian Amazon (Caspar and Rodrigues 1957; Alves 2004; Singerman 2018a,b). We focus in particular on two partially homophonous verbal suffixes: *-psẽ/-pnẽ/-psira*, a RESULTATIVE (Nedjalkov and Jaxontov 1988; Nedjalkov 2001) that agrees with the subject in both number and physical posture, and *-pnẽ/-psira*, a non-witnessed EVIDENTIAL (Chafe and Nichols 1986; Aikhenvald 2004; Brugman and Macaulay 2015) that agrees with the subject in number only. Though the two suffixes are historically related, they can be distinguished from one another on a wide range of synchronic syntactic diagnostics (Singerman 2018a,c). This talk asks: how can we explain the differences between the evidential and resultative in a reader-friendly and accessible way, given that the relevant syntactic diagnostics build upon the configurational structure of the Tuparí clause? More generally: how can reference grammars incorporate the insights of formal syntactic analyses without invoking the representational or theoretical commitments that underlie those analyses?

We propose a two-part solution. First, syntactic diagnostics should be recast in terms of surface order whenever possible. Singerman (2018a,c) analyzes the resultative as attaching at a lower syntactic height than the evidential. While that analysis serves to capture important generalizations about Tuparí clause structure, the main result can be stated in linear terms: the resultative surfaces toward the predicate's left edge, whereas the evidential surfaces towards its right edge. Although this linear restatement does not reflect the organization of the Tuparí clause into distinct layers, it can be understood by readers of various theoretical persuasions and is likely the better choice for a reference grammar.

Second, those diagnostics which cannot be stated in linear terms alone should be accompanied by a clear explanation of the author's analytic assumptions. As concerns the two suffixes under discussion here, only the resultative can occur inside of non-finite nominalizers such as purposive *-tenã*. This is because the evidential occupies a high syntactic position, above the maximum height of *-tenã* and other nominalizers. The resultative, on the other hand, sits beneath the nominalizers' maximum height. This diagnostic makes crucial reference to the attachment site of different morphemes in hierarchical structure (see Baker and Vinokurova 2009, among many others) and cannot be recast solely in terms of linear order. But as this diagnostic captures an important empirical generalization, we believe that it deserves inclusion – with extensive explanation and exemplification – in a comprehensive reference grammar of the Tuparí language.

All data in this talk come from our fieldwork, which has included over nine months of on-site immersion in the Tuparí-speaking community in the Brazilian state of Rondônia.

References

- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.
- Alves, Poliana Maria. 2004. O léxico do Tuparí: proposta de um dicionário bilíngüe [The lexicon of Tuparí: proposal for a bilingual dictionary]. Doctoral Dissertation, Universidade Estadual Paulista, Araraquara.
- Baker, Mark C., and Nadya Vinokurova. 2009. On agent nominalizations and why they are not like event nominalizations. *Language* 85(3):517–556.
- Borsley, Robert D., and Kersti Börjars, eds. 2011. *Non-transformational syntax: formal and explicit models of grammar*. Chichester, West Sussex & Malden, Massachusetts: Wiley-Blackwell.
- Brugman, Claudia M., and Monica Macaulay. 2015. Characterizing evidentiality. *Linguistic Typology* 19(2):201–237.
- Caspar, Franz, and Aryon Dall’Igna Rodrigues. 1957. *Versuch einer Grammatik der Tuparí-Sprache*. Unpublished manuscript.
- Chafe, Wallace, and Johanna Nichols, eds. 1986. *Evidentiality: The linguistic coding of epistemology*. Norwood, New Jersey: Ablex.
- Heine, Bernd, and Heiko Narrog, eds. 2015. *The Oxford Handbook of Linguistic Analysis, 2nd edition*. Oxford: Oxford University Press.
- LaPolla, Randy, and Dory Poa. 2006. On describing word order. In *Catching language: the standing challenge of grammar writing*, eds. Felix K. Ameka, Alan Dench, and Nicholas Evans, 269–295. Berlin; New York: Mouton de Gruyter.
- Mithun, Marianne. 2007. Grammars and the community. In *Perspectives on grammar writing*, eds. Thomas E. Payne and David J. Weber, 45–69. Amsterdam: John Benjamins Publishing.
- Nedjalkov, Vladimir P. 2001. Resultative constructions. In *Language typology and language universals / Sprachtypologie und sprachliche Universalien / La typologie des langues et les universaux linguistiques. Volume 2*, eds. Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible, 928–940. Berlin; New York: Walter de Gruyter.
- Nedjalkov, Vladimir P., and Sergej Je. Jaxontov. 1988. The typology of resultative constructions. In *Typology of resultative constructions*, ed. Vladimir P. Nedjalkov, 3–62. Amsterdam: John Benjamins Publishing.
- Noonan, Michael. 2007. Grammar writing for a grammar-reading audience. In *Perspectives on grammar writing*, eds. David Weber and Thomas Edward Payne, 113–126. Amsterdam: John Benjamins Publishing.
- Rice, Keren. 2006. Let the language tell its story? The role of linguistic theory in writing grammars. In *Catching language: the standing challenge of grammar writing*, eds. Felix K. Ameka, Alan Dench, and Nicholas Evans, 235–268. Berlin; New York: Mouton de Gruyter.
- Rice, Keren. 2007. A typology of good grammars. In *Perspectives on grammar writing*, eds. David Weber and Thomas Edward Payne, 143–171. Amsterdam: John Benjamins Publishing.
- Singerman, Adam Roth. 2018a. The morphosyntax of Tuparí, a Tupían language of the Brazilian Amazon. Doctoral Dissertation, University of Chicago.
- Singerman, Adam Roth. 2018b. Negation as an exclusively nominal category. *Language* 94(2):432–67.
- Singerman, Adam Roth. 2018c. On evidentiality and resultativity in Tuparí. Under peer review.

Description and sociolinguistic function of the secret language of Fedwet -least studied and endangered variety (Etaferahu Tessema, Addis Ababa University)

This research focuses on the social variety used by groups called Fedwet of the Gurage women found in Southern Ethiopia. The term Gurage has been used to designate Semitic-speaking groups surrounded by Cushitic speaking people in the southwest of Addis Ababa, the capital of Ethiopia. The etymology of the name Gurage is still vague. It may be connected to Gura (Hetzorn, 1972:6-7; Hetzron,1977). Leslau (1992:298) also states that Gurage is the most interesting region in terms of the linguistic point of view; the area coverage is relatively small. Gurage is a multidialectal language which appears in diverse varieties of which Fedwet is one of the minority forms.

Fedwet are group of females who were followers of *damam^wit* 'female cult' traditional belief. These groups are part of Gurage specifically from the Chaha clan. They used their own social variety which is not used by males. The social variety took its name, Fedwet, after the Fedwet girls who use it. Although the Fedwet are part of the Gurage society and speak one of the "regular" Gurage languages, they are marginalized groups.

Their variety is also one of the least studied ones. There is no literature on this social variety except Leslau (1964) since he has identified some word lists of the social variety without analyzing its linguistic structure and the sociolinguistic function.

Fedwet speakers acquire the social variety more informally that is they learn from the community and friends who are/were Fedwet without much unconcealed instruction. Thus, the Fedwet social variety is not considered as a means of communication and often stigmatized by the society. The society stigmatizes and doesn't support the use of the Fedwet social variety because they attach the social variety with the traditional belief called *damam^wit*, which is practiced by members of the Fedwet.

The research applied a qualitative method to collect and analyze data. The data were collected through elicitation and in-depth interview. Participants of the research were female members of the Fedwet social group of the Gurage society.

The conceptual framework of the research based on Combination of descriptive linguistics (for gathering and description of linguistic data), comparative linguistics (Fedwet and Fuga social varieties compared with Chaha), and sociolinguistics (function of the social varieties) The result shows that, the group members of Fedwet do not used a different 'language' rather they had fabricated their secret language from their first language Chaha through different modification. The main areas of divergence are the deformation of the morphophonological patterns of their first language Chaha, like substitution, reduplication, insertion, affixation, and syllable manipulation. The other divergence is lexical manipulation, like borrowing, coinage, onomatopieic expression and lexical change also observed some of the terms are changed by other new terms. The other divergence that was used by this group is on Semantic, semantic shift and semantic extension.

There are different reasons that force the females to use a different variety instead of their basic language, Chaha. For instance, the social position of females that is given by the society has a great role to change their ordinary language in different instances. The sociocultural situation and attitude of the society are important to understand women's social position in Gurage. Ethiopia in general and Gurage in particular involve a male dominated sociocultural system. These situations of low societal position push them (Gurage women) to involve different traditional beliefs in addition to their systematized religion.

Hetzron, Robert. 1972. "Ethiopian Semitic: Studies in Classification." *Manchester: Manchester University Press*.

Hetzron, Robert. 1977. *The Gunnän-Gurage Languages*. Napoli: Istituto Orientale di Napoli.

Leslau, Wolf. 1964. "An Ethiopian Argot of a Gurage Secret Society." In *Ethiopian Argots*, 15–22. Paris: Mouton & Co. London Hague.

———. 1992. "Gurage Studies: Collected Article." *Wiesbaden: Harrassowitz*.

Competing motion-cum-purpose strategies in Northern Selkup: a corpus study

S. Toldova (NRU "Higher school of economics", Moscow)

M. Brykina (University of Hamburg; Lomonosov Moscow State University)

Our study follows an appeal from the typology of purpose clauses to perform corpus research on individual languages in order to reveal “principled patterns of usage preferences” [1: 203]. Since we deal with an endangered language with almost no written tradition, the corpus size is not large, however it still allows to make some useful observations.

Northern Selkup is an endangered language from the Southern Samoyedic branch of the Uralic language family. Our study is based on the Selkup corpus developed at Hamburg University, which contains folklore and narrative texts gathered by A.I. Kuzmina in 1960-1970-s [2].

Predicative purpose in Selkup is normally expressed by infinitive constructions. Infinitive is also used with intransitive verbs of motion. However, in this motion-cum-purpose context converbs are found as well. Though no strict constraints on the occurrence of these complement types can be drawn from the corpus data, the question remains whether it is possible to describe the difference in their distribution.

Our claim is that the answer could be suggested within the typological approach. One of the main questions concerning subordinate constructions, including purpose clauses, is the question of syntactic boundedness [3, 1]. Though both infinitival and converb constructions demonstrate quite high degrees of unity, their behavior differs concerning two main parameters of linkage: semantic unity and expression of arguments.

First, though both verb forms have no independent TAM features in Selkup, the infinitival construction has more ‘autonomous’ semantics. The infinitive form presupposes a potential non-realised event, which goes in hand with the cross-linguistic definition of purpose in [1]: the purpose constructions contain an intention component. On the contrary, converb forms in their primary usage (in coordinate constructions) have their TAM interpretation calculated purely on the basis of the TAM and the semantics of the matrix verb. Thus, it is the motion-cum-purpose construction which forces the potential interpretation of the converb form.

Second, while the infinitives have a wide distribution in purpose constructions (cooccurring with matrix verbs of different semantic classes), converbs expressing purpose are mostly limited to the context of motion verbs. Besides, the semantics of the subordinate verb in the two purpose constructions differ. While infinitives tend to occur with accomplishments (1), the converb form is used mainly with activities (2).

Third, while infinitival constructions may be expanded clauses with overtly expressed non-subject arguments or adjuncts (1), there is a tendency for converb clauses to have no complements or adjuncts, except for (often non-referential) DOs (2).

Thus, the difference between the two constructions can be expressed in terms of the typological scale of clause unity. Converb clauses demonstrate higher degree of unity in motion-cum-purpose constructions than the infinitival ones. While the infinitival clauses have a broader distribution in purpose constructions, the converb is a specific form for motion-cum-purpose context and builds a non-compositional construction which enforces the purpose interpretation.

Examples

(1) Tāpi-t nō:kir timn’ā-si-k mora-t tō: na sāri qorqi-p qāt-qa qən-nō:-tit.
 he-GEN three brother-CRC-PL sea-GEN behind this white bear-ACC kill-INF go.away-CO-3PL
Three of his brothers came across the sea to kill this white bear.

(2) Mat mač’ä qon’i-šä-k su:ri-l’-l’ä.
 I to.the.forest go.away-PST-1SG.S wild.animal-VBLZ-CVB
I went to the forest to hunt.

References

- [1] Schmidtke-Bode, K. (2009). *A typology of purpose clauses* (Vol. 88). John Benjamins Publishing.
- [2] The Project „*Grammatical Descriptions, Corpora and Language Technology for Indigenous Northern Eurasian Languages*“: <https://inel.corpora.uni-hamburg.de>
- [3] Van Valin, R. D. (2005). *Exploring the syntax-semantics interface*. Cambridge University Press.

Copulas in Kriyol: approaching a complex case of intralinguistic variation

Chiara Truppi

Universidade de Lisboa

The present paper focuses on copulas in Guinea-Bissau Creole, or simply Kriyol, a Portuguese-based creole language belonging to the group of Upper Guinea Creoles. The paper will describe the system of copulas of Kriyol and approach the variation observed in the expression of predication.

This language has a rich copular system consisting of both verbal (*sedu* ‘to be’, *sta* ‘to be, to stay’, and the past copula (*y)era* ‘was/were’) and non-verbal items (*i* > 3SG subject pronoun *i*, and \emptyset - zero copula). A certain general pattern in the selection of the copula is clearly recognizable: it obeys to criteria, which seem to be shared crosslinguistically, i.e. the TMA-marking of the sentence, and the predicate type. It seems, thus, that each copula is specialized for a certain function. For example, *i* is used in present perfective clauses for attributing non-temporary (inherent) properties to the subject (1a). In the past, the suppletive copula (*y)era* is selected; it may be followed by the past marker *ba* (2a).

This scenario is complicated by the high degree of variation that we observe in this language. For example, we find sometimes \emptyset (1a) or *sedu* (1b) in present-tense copular clauses.

Moreover, there are two further strategies for expressing the past, which use the past marker *ba*: i) *i* (or \emptyset) + COMPLEMENT + *ba* (2b), and ii) *sedu* + *ba* (2c).

- 1) a. *Kil omis (i) piskadur(is).*
 DEM man-PL *i*/ \emptyset fisherman(-PL)
 ‘Those men are fishermen.’
- b. *N sedu monitor.*
 1SG be coach
 ‘I am a [tennis] coach.’
- 2) a. *Kasa ku kai yera (ba) bedju dimas.*
 house REL fall.down be.PST PST old too.much
 ‘The house that fell down was very old.’
- b. *Sedu jornalista i ka nha opson ba nan.*
 Be journalist *i* NEG my option PST indeed
 ‘To be a journalist was not my option.’
- c. [...] *pabia i sedu ba purtugis djagasidu balanta.*
 because 3SG.CL be PST Portuguese mix-PASS Balanta
 ‘[...] because it was Portuguese mixed with Balanta.’

The present paper aims at inserting the description of copulas and copular structures of Kriyol into a broader perspective, in order to deal with the intralinguistic variation found in this language. In more detail, the diachronic perspective will be taken into account: the present-day situation of the copula paradigm in Kriyol is the result of different moments and stages in the emergence of copulas in this language (e.g. the past structure in (2b) is recognized as more ancient by native speakers). Furthermore, possible influences from other languages will be dealt with: both superstrate (Portuguese) and substrate (certain Mande and Atlantic languages such as Mandinka and Wolof) have contributed to the formation of Kriyol and, to a certain extent, to the shaping of its copular system. Finally, we will discuss the importance of constructing a corpus of Kriyol, in order to have a more accurate picture of the frequency of occurrence of each copular item and to notice how regular (possible) semantic differences are.

References

- Ichinose, Atsushi. 1993. Evolução da expressão equacional no Kiriol da Guiné-Bissau. *PAPIA* 2(2). 23-31.
- Kihm, Alain. 1994. *Kriyol Syntax: the Portuguese-based Creole Language of Guinea-Bissau*. Amsterdam: John Benjamins.
- Kihm, Alain. 2007. The two faces of creole grammar and their implications for the origin of complex language. In Regine Eckardt & Gerhard Jäger & Tonjes Veenstra (eds.), *Variation, selection, development: Probing the evolutionary model of language change*, 253-305. Berlin: Mouton de Gruyter.
- Lyons, J. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- Peck, Stephen M., Jr. 1988. *Tense, aspect and Mood in Guinea-Casamance Portuguese Creole*. PhD dissertation, University of California.
- Stassen, Leon. 1997. *Intransitive Predication*. Oxford: Oxford University Press.
- Truppi, Chiara. 2018. The copula cycle in Kriyol, *Diachronic Generative Syntax Conference (DiGS20)*, University of York, UK, June 18-21, 2018. Poster presentation.
- Truppi, Chiara (in submission). Instances of *be* in and around Guinea-Bissau: Predication in Kriyol and an overview of copulas in Upper Guinea Creoles and their substrate.

On the challenges of letting a language tell its own story: evidence from the Amazon

Rosa Vallejos
University of New Mexico

This talk deals with some of the challenges field linguists face during the grammar writing process, particularly in the context of working with little-known, underdescribed languages. In addition to collecting and processing substantial amounts of data, there are many decisions one must take. Who is the intended audience? If the language is endangered, should the grammar include information with enough detail to be useful for language development projects? Should it easily allow for comparative work between languages of the family? Should the analysis be placed in the context of crosslinguistic patterns and existing typologies? Should the grammar address current questions on linguistic theory? There is certain consensus that grammar writers should allow a language to tell its own story (see, for instance, Rice 2005, 2006; Dixon 2009). In sum, “each language demands its own strategy of presentation” (Rice 2005: 400). Yet the pedagogical uses of grammars cannot be overlooked. Reference grammars are often used as tools and sources of data to develop an understanding of the diversity of grammatical constructions in the world’s languages. As every grammar is organized differently, and frequently contains idiosyncratic terminology, obtaining information from them is an acquired skill (Haspelmath 2010; Croft 2016).

This talk takes gender and possession in Kukama-Kukamiria (Peruvian Amazon, henceforth KK) as cases in point for discussing methodological and theoretical issues when presenting empirical facts that reveal multiple connections. This language does not display grammatical gender, it instead has a system that indexicalizes the gender of the speaker. This genderlect system is manifested primarily in the pronominal system, the demonstrative modifiers, and certain discourse connectors and interjections. Overall, KK displays one of the most pervasive gender indexicality systems amongst the languages documented to encode genderlects (Fleming 2012; Rose 2015). It has one of the highest degrees of indexicality in terms of both the number of domains in which it is codified, and the frequency of appearance of gender indexicals in spontaneous, communicative interactions. In sum, gender marking affects all areas of the grammar.

Regarding possession, KK does not have lexical verbs such as ‘have’, ‘belong,’ a copula, or a grammatical construction to predicate ownership. Although the language does not have a dedicated possessive construction, possession can be inferred from other constructions, including the three below:

- (1) *ajan wayna mimira-yara*
this woman woman’son-owner
‘This woman has a son’ (Lit. ‘This woman is son-owner’) [female speech]
- (2) *emete-taka ajan wayna iki*
exist-MOD this woman chili
‘This woman might have chili’ (Lit. ‘This woman’s chili might exist’) [female speech]
- (3) *ta yuwama=ka emete eran kaitsuma*
1SG.M daughter.law=LOC exist good yucca.beer
‘My daughter in law has good yucca beer’
(Lit. There is good yucca beer at my daughter in law) [male speech]

Examples in (1) - (3) provide evidence of a conceptual link between categorization, location, existence, and possession. The example in (1) is a predicate nominal construction. It consists of two juxtaposed NPs; the possessor is encoded in the first NP, the possessed in the second NP. The second NP includes the clitic *-yara* ‘owner.’ The example in (2) illustrates an existential construction in which the possessor and possessed elements are expressed within an NP; importantly, the existence of the NP’s referent is being predicated. The example in (3) is a combination of the existential and a locative construction. Here, the new referent’s location is specified; the possessor is expressed in the locative phrase, the possessed element in the NP.

With these examples in mind, what aspects must be explicitly indicated in the grammar? What is optional? What is “obvious”? I will discuss the strategies employed to present these empirical facts regarding gender and possession in a cohesive way to the widest possible audience.

“An helpful guideline based on areal typology for describing Southeast Asian languages”

Alice Vittrant¹ & Justin Watkins²

alice.vittrant@cnrs.fr

jw2@soas.ac.uk

¹Aix-Marseille University (AMU)/CNRS-DDL

²School of Oriental and African Studies (SOAS), London

Thomason and Kaufman’s 1988 book *Language contact, creolization, and genetic linguistics* had a stimulating effect on the fields of comparative and descriptive linguistics and inspired a number of studies on various topics related to language contact: the relationship between typology and language contact; the effect of language contact on a language’s genetically inherited characteristics, and work on mixed and endangered languages.

Within this landscape, our proposal lies at the crossroads of these themes, with the following aim : to provide a frame for describing endangered languages of a particular areal, i.e. Southeast Asia (henceforth SEA). SEA is known by linguists involved in areal linguistics as being a ‘contact superposition zone’ (Koptjevskaja-Tamm & Wälchli 2001), a zone of contact-induced structural convergence’ (Bisang 2006) or simply a *Sprachbund* or linguistic area (Matisoff 1986, Enfield 2005, Vittrant 2010). It means that both languages with a close genetic relationship and languages from different language families are found in this area. However, despite lacking a common origin, these languages had surprising structural similarities, apparently acquired in part through contact with structural linguistic features that they did not originally possess.

In recent decades a number of definitions of a linguistic area or *Sprachbund* have been proposed (see Campbell 2006; Stolz 2006:33), all of which aim to describe the phenomena of linguistic convergence, common innovations or common retentions, or to identify the specific properties which set a *Sprachbund* apart from other language contact situations. However, despite numerous attempts to define the concept precisely, a consensus emerges on the impossibility of identifying universal criteria.

The problems encountered in seeking to define a linguistic area, however, do not detract from the relevance of studies of changes induced by contact between the languages spoken within a particular geographical area, i.e. areas of linguistic convergence. In the case of Southeast Asia, a region characterized by the presence of five language families and several millennia of contact between the area’s linguistic communities, an areal approach is fruitful both for the description of undescribed languages and for typological studies.

Based on previous studies that draw together features that cut across the genealogical phyla in many domains, inspired by existing typological studies of specific grammatical phenomena such as Kahrel and van den Berg’s (1994) work on negation, or cross-linguistic studies such as Aikhenvald & Dixon (2003), (2006), Zúñiga & Kittilä (2010), and Kopecka & Narasimhan (2012) inter al., we proposed a guideline (largely illustrated) to linguists who want to describe languages of SEA. This guideline suggests to follow a particular structure and typological approach. Thus, adopting a broadly similar organization and structure and using similar terminology will allow researchers to do cross-comparisons. It will also facilitate typological studies, i.e. studies on specific linguistic phenomena across languages. Lastly, it may show that the language studied is on the hedge of the sprachbund lacking some of the MSEA features... and exhibiting features that are representative of the adjacent areas.

The guideline also suggests to add a glossed text to any language description, offering a glimpse of the language used in more natural context, highlighting also the increasing role of Corpora in descriptive linguistics.

By providing this guideline, we hope to inspire further work on the many languages of Southeast Asia which have yet to be described.

By providing this guideline, we will also help to see how prototypical is the language described

References :

- Aikhenvald, A.Y. & Dixon R.M.W. (eds). 2003. *Studies in Evidentiality*. Amsterdam/Philadelphia: Johns Benjamins
- Aikhenvald, A.Y. & Dixon R.M.W. (eds). 2006. *Serial Verb Constructions: A Cross-Linguistic Typology*. New York : Oxford University Press.
- Bisang, Walter. 2006. Contact-induced convergence: Typology and areality. In: Brown, Keith. (ed.), *Encyclopedia of Language and Linguistics*, volume 3. Oxford: Elsevier. 88–101.
- Campbell, Lyle. 2006. Areal linguistics: A closer scrutiny. In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic areas: Convergence in historical and typological perspective*, 1–31. Hampshire: Palgrave Macmillan.
- Enfield Nick J. 2005. Areal linguistics and Mainland Southeast Asia, *Annual Review Anthropology*, 34. 181–206.
- Kahrel, Peter & R. Van Den Berg (eds.). 1994. *Typological Studies In Negation*, Amsterdam: John Benjamins.
- Kopecka, Anetta & Bhuvana Narasimhan. 2012. *Eventes of Putting and Taking. A crosslinguistic perspective*. Amsterdam: John Benjamins.
- Koptjevskaja-Tamm Maria & Bernhard Wälchli. 2001. The Circum-Baltic languages: an areal-typological approach. In Östen Dahl & Maria Koptjevskaja-Tamm (eds), *The Circum-Baltic languages: typology and contact*. (Vol. 2) Grammar and typology. Amsterdam: John Benjamins. 615–750.
- Matisoff, James A. 1986. Linguistic diversity and language contact. In John McKinnon & Wanat Bhruksasri (eds.), *Highlanders of Thailand*. Singapore: Oxford University Press. 56–86.
- Stolz, Thomas. 2006. All or nothing. In Matras Yaron, April McMahon & Nigel Vincent (eds), *Linguistic Areas: convergence in Historical and Typological Perspective*. Hampshire: Palgrave. 32–50.
- Vittrant Alice. 2010. Aire linguistique Asie du Sud-Est continentale: le birman en fait-il partie ? *Moussons* 16. 7–38.
- Zúñiga, Fernando & Seppo Kittilä. 2010. *Benefactives and Malefactives. Typological perspectives and case studies*. Amsterdam: John Benjamins.

Towards a corpus-based grammar of Upper Lozva Mansi: diachrony and variation¹

Daria Zhornik^{1,2}, Sophie Pokrovskaya¹, Lev Kozlov⁴, Vladimir Plungian^{1,2,3}

¹Lomonosov Moscow State University

²Institute of Linguistics RAS

³Vinogradov Russian Language Institute RAS

⁴National Research University Higher School of Economics

Mansi (<Ob-Ugric <Finno-Ugric <Uralic) is an endangered and underdescribed language spoken in Western Siberia (2010 census: 938 speakers). Most people who still use Mansi in everyday communication reside in the Northern part of the Sverdlovsk Oblast and speak the Upper Lozva dialect (~100 speakers). Our project is aimed at the documentation of the Upper Lozva dialect which, in time, inevitably leads to writing a grammar. In our talk, we will discuss challenges which arise during our work and propose possible solutions.

Our needs are as follows:

- produce a grammar of an endangered Mansi dialect which is easy to incorporate into scientific discourse
- that is, write a grammar from a typological point of view
- describe the Upper Lozva dialect as chronologically the “latest” version of the Mansi language
- at the same time, we have to describe it as a regional dialect with its own traits.

A grammatical description should be integrated into the typological background to enable efficient comparison with other languages as well as to avoid inadequate interpretation of grammatical phenomena found elsewhere in the world. To describe a dialect from a typological point of view, we need to outline the protolanguage system, investigate paths of its divergence and describe the way it developed into the dialect in question.

To write such a grammar, we need as much natural data as possible: elicitation questionnaires, text corpora, multimedia data. The corpus of Mansi texts that is currently being created by the members of our project has wide dialectal and temporal coverage (all Mansi varieties, texts recorded in 1840-2018). A particularly valuable tool is the Upper Lozva multimedia subcorpus which consists of data gathered in the field in 2017-2018 and features audio files, which are necessary not only for phonetic studies – freely available data may be reanalyzed by the linguistic community. While gathering data in the field, we apply the method proposed by A. E. Kibrik (see e.g. [Kibrik 2010]), which assumes approaching the language as a *tabula rasa* and creating a new grammar from scratch without getting biased by previously available descriptions.

Synchronic sketches of separate dialects are usually unable to successfully incorporate linguistic data into scientific discourse due to the absence of diachronic components and disregard of dialectal variation. Researchers apply different methods while writing grammars, and these diverse approaches need to be linked somehow. We believe that this link may be found while regarding the language as a system which is extended both in time and space. While comparing various dialects or varieties pertaining to different time periods we no longer need to aim at applying a certain method as we only describe pure linguistic data.

The classical format of “historical grammars” suits us well, as it implicitly models the language as a result of a set of transformations of a proto-system. This is important, as diachrony often mirrors the synchronic variation found in the language (see e.g. [Kuryłowicz 1962]).

¹ The current research is supported by the RFBR grant №18-012-00833

We restrain from creating a formal model as our main concern is not finding the universal features of the language but to highlight special Mansi traits. In our case, a large part of the existing variation between Mansi idiolects lies in the surface structure, that is, phonetics, which may vary even in the speech of a single speaker. A formal model is not capable of capturing such fine-grained types of variation, which are of utmost importance for us. Studying such patterns of variations enable us to understand where the system has undergone changes, and, what is even more significant, where we should expect further change.

We use the traditional scheme of grammar writing; however, each section should be divided into subsections describing the situation in the protolanguage, in other Mansi dialects and, finally, in the Upper Lozva dialect. This method of description should be universally useful, as each language consists of a certain number of more or less closely related varieties.

Another problem that we face is the role of native speakers in writing grammars (see the discussion on integrating members of the local community into linguistic work in [Jones, Mooney 2017]). The only full-fledged grammar of Northern Mansi was written by a native speaker [Rombandeeva 1973] and has numerous setbacks, e.g. most examples given in the grammar have been produced by the author herself and, practically, only reflect her own idiolect [Burlak, Starostin 2005]. As a speaker's knowledge of their own language is intuitive, it is difficult to approach language facts objectively, which leads to a strongly biased language description. On the other hand, a linguist alone is not always capable of creating an in-depth description without having advanced knowledge of the language in question (which is often next to impossible to acquire, as field trips are only temporary and do not always provide necessary conditions for successful language learning). Thus, we attempt to unify the analytical and empirical approaches [Kant 2007] by bringing together both linguists and native speakers at all stages of our work.

References

1. Burlak, Starostin 2005 – Burlak S., Starostin S. *Sravnitel'no-istoricheskije jazykoznanije*. M.: Academia, 2005.
2. Jones, Mooney 2017 – Jones M., Mooney D. (eds.) *Creating orthographies for endangered languages*. Cambridge: Cambridge University Press, 2017.
3. Kant 2007 – Kant I. *Kritika chistogo razuma*. M.: Eksmo, 2007.
4. Kibrik 2010 – Kibrik A. E. *Ekspedicionnyje istorii 3: prodolzhenije dagestanskih sopostavitel'nykh shtudij (1974-1977) // K 70-letiju V. A. Vinogradova*. M., 2010.
5. Kuryłowicz 1962 – Kuryłowicz J. *Ocherki po lingvistike*. M.: Izdatel'stvo inostranoj literatury, 1962.
6. Rombandeeva 1973 – Rombandeeva E. I. *Man'sijskij (vogul'skij) jazyk*. M.: Nauka. 1973.