Sensory landscapes in motion: on the directional encoding of perceptual events in English and Russian

Among studies on sensory language, there has been an increasing interest in spatial encoding of visual experiences. For instance, it has been observed that visual events are expressed via directional forms (1) in Indo-European (Cappelle, 2020), Uralic (Huumo, 2010), Tai–Kadai (Takahashi, 2000), Austroasiatic (Wnuk, 2022) and Mayan (Craig, 1993) languages.

(1) *Hunter **looked into his eyes** reassuringly.*

Meanwhile, spatial encoding and conceptualization of senses in inter-modal perspective, i.e. comparative research on different sensory modalities, remain largely understudied. Hence, the present paper investigates directional encoding of visual, auditory and olfactory events in English and Russian.

To explore the spatial framing of perceptual events, I adopt the notion of *sensory path* and assume that such events involve "two entities, the Experiencer and the Experienced, and [...] something intangible moving in a straight path between them in one direction or the other" (Talmy, 2000: 115). The paper aims at analyzing (i) distribution of directionality patterns (Experiencer → Experienced vs. Experienced → Experiencer) across three sensory modalities and (ii) lexicalization patterns of sensory path (i.e. regularities in meaning-to-form associations) in two languages.

For this study, I built a parallel corpus of contemporary texts written in English and Russian and belonging to various literary genres (e.g. crime fiction, fantasy, science fiction). In contrast to previous studies on sensory paths, I extracted all constructions describing perceptual events that involved directional encoding, including perception (1), self-motion ((2); (3)) and caused motion (4) predicates.

(2) *A soft rustling seemed **to be coming from up ahead.***
(3) *A sudden smell **reached his nose**.*
(4) *The thin man **threw Artyom an angry look.***

Regarding lexicalisation patterns, both languages rely on satellites (prefixes in Russian and particles in English) and PPs in the expression of path across different sensory modalities. However, there is a difference in the use of verbs describing the motion of sounds: Russian extensively uses manner-of-motion verbs (e.g. *letet'* 'fly'), while in the English sub-corpus we mostly find the deictic verb *come* (2), with manner being unspecified. This suggests that the typological similarities these two languages present in encoding human locomotion (Slobin, 2005), may not directly map onto the expression of motion in the domain of perception.

As for directionality patterns, in both languages visual events are predominantly conceptualized following the Experiencer → Experienced path ((1), (4)). On the other hand, auditory and olfactory events mostly follow the Experienced → Experiencer pattern ((2); (3)).

I hypothesize that this variation may be explained by the different degree of control over the stimuli in vision as compared to hearing and olfaction. The subject of a visual experience can select a stimuli by moving their eyes, whereas the subject of an auditory or olfactory experience is less agentive in this regard. Therefore, I argue that the orientation of a sensory path varies across sensory modalities. Further research could elucidate if the aforementioned folk model of perception applies across typologically varied languages.

References

Cappelle, B. (2020). Looking into visual motion expressions in Dutch, English, and French: How languages stick to well-trodden typological paths. In Y. Matsumoto & K. Kawachi (eds.), *Broader Perspectives on Motion Event Descriptions*, 235-280. Amsterdam/Philadelphia: John Benjamins.

Craig C., 1993, Jakaltek directionals: their meaning and discourse function, *Languages of the World 2*, Lincom Europa, p. 23-36.

Huumo, T. (2010). Is perception a directional relationship? On directionality and its motivation in Finnish expressions of sensory perception. *Linguistics*, *48*(1), 49-97.

Slobin, D. (2005). Relating narrative events in translation. In D. Ravid & H. Shyldkrot (eds.), *Perspectives on language and language development: Essays in honor of Ruth A. Berman*, 115-129. New York: Springer

Takahashi, K. (2000). *Expressions of emanation fictive motion in Thai*. https://scholar.google.com/scholar_lookup?title=Expressions+of+emanation+fictive+motion+e vents+in+Thai&author=Takahashi+Kiyoko&publication+year=2000

Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Wnuk, E. (2022). Ways of looking: Lexicalizing visual paths in verbs. *Journal of Linguistics*, *58*(1), 157-202.

Corpus evidence of optional ergative marking in Ika (Chibchan, Colombia)

Jana Bajorat

January 19, 2024

Abstract

In Ika, an understudied Chibchan language spoken by approx. 25,700 people (number from 2018, DANE (2021)) in Colombia, ergative case marking is optional (see, e.g., McGregor, 2010; Riesberg, 2018): In basic canonical clauses, subjects and objects are unmarked (SOV-order). Only 3rd person transitive subjects can be case marked under certain circumstances with the ergative marker $=se'$. Previous research has attributed multifaceted functions to this marker: Levinsohn & Tracy (1977: 7-8) describe the marker as an indicator of 'change of [thematic] role', and also Frank (1985) explains its presence through discourse-pragmatic factors such as marking of subjects that are less given than the object, and 'reintroduced' or 'unexpected' agents. Yet, these characterisations are fuzzy and in need of updating to current terminology (see, e.g., Krifka & Musan, 2012). Other accounts ascribe the ergative marker a disambiguating function in transitive clauses: Landaburu (2000: 744) characterises it as a 'non-object' marker in the context of topicalisation, and Frank (1990) claims that $=se'$ identifies the subject in clauses with non-canonical argument order, i.e. $S_{TRANS}V$, OSV, SVO, or OVS. These analyses, however, do not explain the occurrences of the ergative marker in canonical argument order. Own elicited and experimental data reveal that only a combination of both accounts offers a satisfactory explanation: first, the ergative marker is obligatory in non-canonical argument orders ($S_{TRANS}V$, OSV, SVO, OSV) in order to be able to disambiguate subject and object (cf. example 1). Second, the marker is optional in canonical SOV-order since argument order serves as primary mechanism for argument distinction and makes morphological marking functionally redundant (cf. example 2 - both versions are grammatically correct and do not differ semantically from each other, although the version with the ergative marking is strongly preferred). Third, information structure, presumably focus or prominence marking (cf. Krifka & Musan, 2012; Heusinger & Schumacher, 2019), is the underlying trigger for the overt ergative marking in canonical argument order, but is, of course, also responsible for the changes in argument order that lead to the obligatory ergative marking.

(1) Pedru Juan pɨs-ɨn   nug-in.
     Pedru Juan hit-IPFV AUX.NONEGO-DECL

     'Pedro is hitting Juan.'   SOV

(2) Juan Pedru pɨs-ɨn   nug-in.
     Juan Pedro hit-IPFV AUX.NONEGO-DECL

     'Juan is hitting Pedro.'   SOV

(3) Juan Pedru=se'   pɨs-ɨn   nug-in.
     Juan Pedru=ERG hit-IPFV AUX.NONEGO-DECL

     'Pedro is hitting Juan.'   OSV

(4) Pedru pɨs-ɨn   nug-in.
     Pedru hit-IPFV AUX.NONEGO-DECL

     'He/she/it is hitting Pedro.'   OV

(5) Pedru=se'   pɨs-ɨn   nug-in.
     Pedru=ERG hit-IPFV AUX.NONEGO-DECL

     'Pedro is hitting him/her/it.' $S_{TRANS}V$

The present study will confirm the above observations (which are based on experimental and elicitation evidence) with corpus data. For this purpose, an Ika spoken language corpus

was created. The corpus consists of 70 narratives recorded mainly in 2022 with 10 different speakers (7 text types per speaker, including traditional story telling and descriptions of house constructions, paths, cooking, and comparison with other indigenous people; total corpus length: 5h 46 min 32 sec). The corpus is (at the moment partly) annotated for syntax, specifically for argument order, ergative marking and transitivity, and also for information structure. In this talk, I will present quantitative results supporting the dependency of overt ergative marking on non-canonical argument order or focus/prominence. I will show examples together with an analysis of the specific information-structural notion that ergative marking signals on the subject in these clauses.

Initial results confirm the above made assumptions for the ergative marking distribution pattern, namely that ergative marking occurs obligatorily in non-canonical argument orders ($S_{TRANS}V$ is the most frequent pattern) and that, in canonical argument orders, the occurrence of the ergative marker signals prominence or focus. This marking is relatively frequent, as it occurs in about a third of SOV-clauses, which themselves represent approx. a fourth of all transitive clauses.

References

DANE. (2021). *Resultados del censo nacional de población y vivienda 2018: Pueblos arhuaco, kankuamo, y wiwa.* Retrieved 19.01.2023, from `Mhttps://www.dane.gov.co/files/investigaciones/boletines/grupos-etnicos/presentacion-pueblos-arhuaco-kankuamo-wiwa-CNPV2018.pdf`

Frank, P. S. (1990). *Ika syntax* (Vol. 1). Dallas, TX and [Arlington]: Summer Institute of Linguistics and University of Texas at Arlington.

Heusinger, K. v., & Schumacher, P. B. (2019). Discourse prominence: Definition and application. *Journal of Pragmatics, 154*, 117–127. doi: M10.1016/j.pragma.2019.07.025

Krifka, M., & Musan, R. (2012). Information structure: Overview and linguistic issues. In M. Krifka & R. Musan (Eds.), *The expression of information structure* (pp. 1–43). Berlin and Boston: Mouton de Gruyter.

Landaburu, J. (2000). La lengua ika. In González de Pérez, María Stella & Rodríguez de Montes, Maria Luisa (Eds.), *Lenguas indígenas de colombia* (pp. 733–748). Santafé de Bogotá: Instituto Caro y Cuervo.

Levinsohn, S. H., & Tracy, H. (1977). Participant reference in ika expository discourse. In R. E. Longacre & F. Woods (Eds.), *Discourse grammar: studies in indigenous languages of colombia, panama, and ecuador* (pp. 3–24). Dallas, Texas: Summer Institute of Linguistics and University of Texas.

McGregor, W. B. (2010). Optional ergative case marking systems in a typological-semiotic perspective. *Lingua, 120*(7), 1610–1636. doi: M10.1016/j.lingua.2009.05.010

Riesberg, S. (2018). Optional ergative, agentivity and discourse prominence – evidence from yali (trans-new guinea). *Linguistic Typology, 22*(1), 17–50. doi: M10.1515/lingty-2018-0002

Verbale Kongruenzmarkierung in europäischen Gebärdensprachen

Kevin Behrens (Universität Bremen)

Gebärdensprachen weisen eine große Bandbreite an Kodifizierungsstrategien für verbale Kongruenz auf. Von besonderem Interesse ist hierbei die Grammatikalisierung des Gebärdenraums, in den Satzargumente an bestimmte Punkte des Raumes platziert werden können. Auf diese kann Rückbezug genommen werden, unter anderem um am Verb das Subjekt und/oder Objekt zu markieren. Von den verschiedenen Verbklassen in den Gebärdensprachen wird an direktionalen Verben diese Markierung direkt vorgenommen, einfache Verben ohne eigene Kongruenzstellen können hierzu in manchen Gebärdensprachen ein Kongruenzauxiliar nehmen, das dieselbe Markierungsstrategie benutzt (Quer et al. 2019: 206). Es wird oft angenommen, dass diese grammatische Raumnutzung wie in (1) universell unter den *Deaf community sign languages* ist.[1]

(1) Deutsche Gebärdensprache (Eichmann et al. 2012: 88)[2]
    MORGEN 1-BESUCHEN-re
    ‚Morgen besuche ich ihn.‘

In meinem Vortrag ziehe ich zwischen 15 europäischen Gebärdensprachen einen Vergleich im Hinblick darauf, wie sie bei Verben die Personenkongruenz markieren. Das beinhaltet zum einen das Ausmaß und die Produktivität der direktionalen Raumnutzung. Denn die Menge an direktionalen Verben unterscheidet sich zwischen den Sprachen, ebenso deren Funktionsumfang. Zum anderen zeige ich auf, welche unterschiedlichen Markierungsstrategien die Gebärdensprachen aufweisen. So hat die Deutsche Gebärdensprache (DGS) neben direktionalen (mono- und bidirektionalen sowie rückwärtsgerichteten) Vollverben ebenfalls Auxiliare, die kongruieren. Das PAM-Auxiliar (Personal Agreement Marker) wird zum Beispiel genutzt, um einfache, nicht-direktionale Verben zur Kongruenzmarkierung zu verhelfen, zu sehen in (2). Dementgegen kennt aber sowohl die Britische als auch die Amerikanische Gebärdensprache keine Kongruenzauxiliare, sehr wohl aber auch direktionale Verben. Teils wurde bei ihnen aber festgestellt, dass sie durch non-manuelle Marker wie Blickrichtung oder Kopfneigung Kongruenz markieren (Pfau et al. 2018: 36).

(2) Deutsche Gebärdensprache (Eichmann et al. 2012: 91)
    SOHN SITZEN-re LESEN. VATER PAM-re BEGEISTERT
    ‚Der Sohn sitzt und liest da. Der Vater ist von ihm begeistert.‘

Es soll gezeigt werden, welche Faktoren die möglichen Unterschiede in der Markierung bedingen können. Zum Beispiel lässt sich ein Zusammenhang zwischen Zeittiefe der Sprachgenese und der grammatischen Raumnutzung feststellen (Pfau et al. 2018: 36). Aber auch die Phonologie spielt eine Rolle, so können körpergebundene Gebärden ihre Ausgangsposition nicht verändern, um mit dem Subjekt zu kongruieren. Durch Sprachwandel wird aber bei immer mehr Gebärden die Körpergebundenheit abgelegt, um aus einem einfachen Verb ein direktionales zu machen, zum Beispiel bei dem Verb FRAGEN. Nur noch in der Bedeutung des Nomens ‚Frage‘ wird die Ausgangsposition am Kinn gebärdet, das Verb ist in diesem Parameter frei variierbar geworden (Papaspyrou et al. 2008: 127). Welche weiteren Rollen Verwandtschaft, Sprachkontakt bzw. Regionalität oder Universalität hierbei im europäischen Raum spielen und wie sich die Gebärdensprachen hier zu den Lautsprachen vergleichen lassen, lege ich in meinem Vortrag dar.

---

[1] Diesen Gemeinschaftsgebärdensprachen stehen *Village sign languages* durch ihre sozio-linguistischen Umstände gegenüber, die interessanterweise keine Raumgrammatikalisierung kennen (Meir et al. 2010).

[2] Die Markierung *1-* steht für den Beginn der Bewegung von BESUCHEN an der Position für die 1. Person, *-re* für die Endposition an einem rechtsgelegenen Punkt im Gebärdenraum zur Markierung einer 3. Person.

Quellen

Eichmann, Hanna, Martje Hansen, Jens Heßmann. 2012. *Handbuch Deutsche Gebärdensprache – Sprachwissenschaftliche und anwendungsbezogene Perspektiven*. Seedorf: Signum Verlag.

Meir, Irit, Wendy Sandler, Carol Padden, Mark Aronoff. 2010. Emerging Sign Languages. In Irit Meir, Wendy Sandler, Carol Padden, Mark Aronoff (Hg.), *Oxford Handbook of Deaf Studies, Language, and Education*, Vol. 2, 268–280. Oxford: Oxford University Press.

Papaspyrou, Chrissostomos, Alexander von Meyenn, Michaela Matthaei, Bettina Hermann. 2008. *Grammatik der Deutschen Gebärdensprache – aus Sicht gehörloser Fachleute*. Seedorf: Signum Verlag.

Pfau, Roland, Martin Salzmann & Markus Steinbach. 2018. The syntax of sign language agreement: Common ingredients, but unusual recipe. *Glossa: a journal of general linguistics* 3(1): 107. 1–46.

Quer, Josep, Carlo Cecchetto, Caterina Donati, Carlo Geraci, Meltem Kelepir, Roland Pfau, Markus Steinbach. 2019. *SignGram Blueprint – A Guide to Sign Language Grammar Writing*. Berlin & Boston: Walter de Gryuter.

How concessive conditionals benefit the semantic map of additivity

In this paper, I propose an improvement to Forker's (2016) semantic map of additivity. This proposal is the result of a cross-linguistic investigation into a type of adverbial clause known as "concessive conditionals". Additive markers (ADD), e.g. German *auch*, express that there is at least one alternative value for their associate, i.e. the linguistic element in their scope (Forker 2016; König 1991). In many languages of the world, additive markers are highly polyfunctional. They may, among other functions, occur in indefinite pronouns and in adverbial clauses such as concessives and concessive conditionals. Following Haspelmath & König (1998), I distinguish three subtypes of concessive conditionals: scalar concessive conditionals or SCCs ('even if …'), alternative concessive conditionals or ACCs ('whether … or …'), and universal concessional conditionals or UCCs ('WH-ever / no matter WH …'). The main difference between concessive conditionals and concessives proper ('although ….') is that the subordinate clause typically evokes a set of hypothetical antecedents in the former, but a single factual value in the latter (Haspelmath & König 1998: 567). The functions most commonly expressed by additives in the languages of the world are shown in Forker's (2016) semantic map in Figure 1.

Figure 1 has a dotted line between the scalar additive and indefinite nodes, indicating a "significant semantic link" between these two functions (Forker 2016: 87). I will show that there is clear cross-linguistic evidence that indefinites should not be connected to scalar additivity, but to the concessive node, more specifically to UCCs. This has gone unnoticed because Forker subsumes concessive conditionals and concessives proper under the same node. The indefinite node, too, is complex, comprising specific, universal, negative, and free-choice indefinites (ibid.: 79). Once disentangled, it becomes clear that indefinites, in particular free-choice indefinites (FCIs), are connected to UCCs. In Wolaytta (Omotic, Ethiopia), for example, both FCIs and UCCs are formed by adding the additive marker *-kka* or *-nne* to a question word or a conditional clause containing a question word, respectively.

Having demonstrating this, I will briefly discuss (a) whether the link between indefinites and the "core" additive function or UCCs should be considered the main one (full vs. dotted line) and (b) the diachrony of this connection, i.e. whether FCIs introduce UCCs or whether FCIs tend to develop out of UCCs. Preliminary results suggest that the latter scenario is more likely (cf. Haspelmath 1997: 159–163).

References

Forker, Diana. 2016. Toward a typology for additive markers. *Lingua* 180, 69–100.

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

Haspelmath, Martin & Ekkehard König. 1998. Concessive conditionals in the languages of Europe. In Johan van der Auwera (ed.), *Adverbial constructions in the languages of Europe*. Berlin: de Gruyter, 563–640.

König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective*. London: Routledge.

Wasaka, Motomichi. 2008. *A descriptive study of the modern Wolaytta language*. Doctoral dissertation, University of Tokyo.
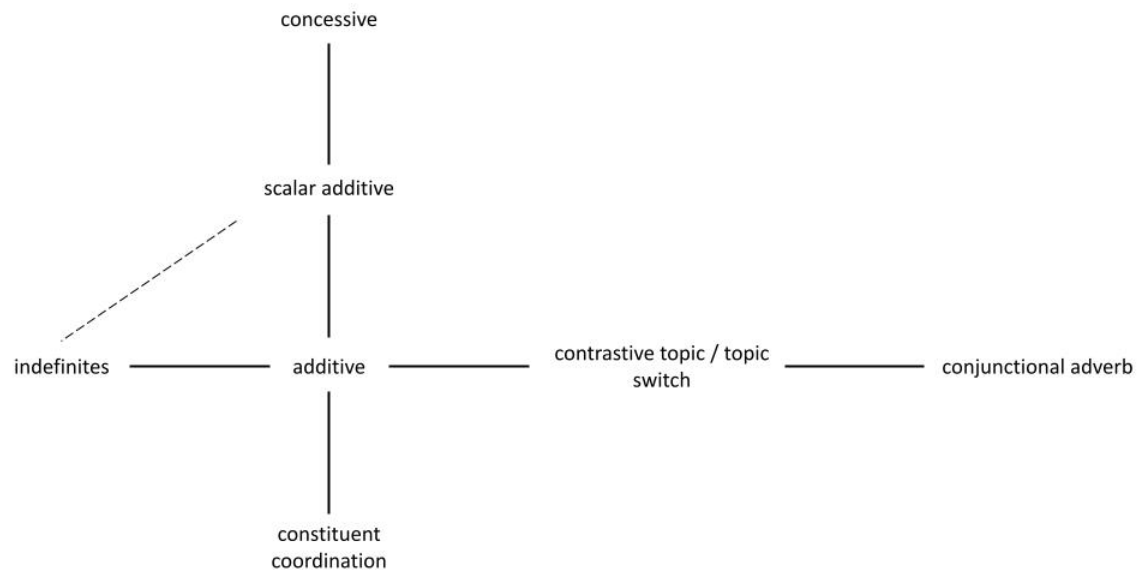
Figure 1: Semantic map of additivity by Forker (2016)

Brunner Jennifer, University of Graz

Directives in Central Pame, Mexico

The only existing work clearly dedicated to imperatives in the Mexican language Central Pame (pbs, cent2145) to date is a typewritten manuscript of three and a half pages. Olson's *Fifty Pame Imperatives* (1951) is a list of fifty imperative forms addressing 2nd persons in singular, dual and plural. His work is the starting point for an extended and revised analysis of directives in this Otopamean language. Olson's concentration on form is widened by a focus on function. The discussion of imperatives is integrated into a larger picture dealing with directives and their place in the verbal system of Central Pame as a whole (Brunner 2023).

The paper addresses canonical (addressing 2nd persons) and non-canonical forms (1st and 3rd persons) of different directive-types – like imperatives, polite requests and prohibitives – as well as number- and object-marking (Hurch 2021, Brunner 2016). A corpus of 272 verbs resp. their canonical imperative forms in singular, dual and plural are the base for a re-formulation of Olson's form-classes. The main data sources are Olson (1951), Gibson (1994), Hurch (2012–2023) and Herce (2023) supplemented by material predominantly relevant for the study of non-canonical forms and negative imperatives like stories (Gibson, Olson & Olson 1963; Gibson 1966) and my own collections (Brunner 2022, 2024).

The results are that Central Pame shows an independent imperative paradigm sensitive to transitivity. Additionally, other ATAM[1]-forms, namely *potential* and *perfective*, come into play in order to express polite nuances and negative commands.

Selected References

Brunner, Jennifer. 2016. *Zur Objektmarkierung im Pame Central und darüber hinaus.* Masterarbeit. Graz: Universität Graz.

Brunner, Jennifer. 2022. *Fragebogen zu negativen Imperativen.* Unv. Ms.

Brunner, Jennifer. 2023. *Von Aktionalität, Temporalität, Aspektualität und Modalität im Zentralpame (šiʔui pame), Mexiko. Dissertation. Graz: Universität Graz.*

Brunner, Jennifer. 2024. *Fragebogen zu Hortativen.* Unv. Ms.

Gibson, Lorna. 1966. The Man Who Abandoned his Wife. A Pame Story. *Tlalocan* 5.2, 169–177.

Gibson, Lorna. 1994. *Verb Paradigms in Pame of Santa María Acapulco, San Luís Potosí*. Eingetragen v. Ramona Millar. Hrsg. v. Doris Bartholomew. Ms. SIL, Language & Culture Archives, https://www.sil.org/system/files/reapdata/14/87/38/14873814102253810718722755643232320826829/pbs _Verbs.pdf (30.03.2024).

Gibson, Lorna, Donald Olson & Anne Olson. 1963. Four Pame Texts. *Tlalocan* 4.2, 125–143.

Herce, Borja. 2023. *VeLePa: a Verbal Lexicon of Pame*. https://osf.io/xhyzm/files/osfstorage (13.03.2024).

Hurch, Bernhard. 2012–2023. *Paradigmen von Verben und Nomen im Pame erhoben mit Felipa Landaverde Durán[,] Alfreda Reyes Botello und Antonia Montero Montero[,] Josefa Hernández de la Cruz, Carlos Ramos Ramos[,] [Juan Martínez Izeaguirre, Rufino Medina Marta] im Rahmen der Arbeit zum Pame-Band des Archivo de Lenguas Indígenas de México*. Unv. Ms. [Stand: Februar 2024]

Hurch, Bernhard. 2021. Sobre el sistema de pronombres de persona en pame. Un estudio comparativo y algunas observaciones sobre la posición del pame meridional. *Estudios de Cultura Otopame* 12, 187–229.

Olson, Donald. 1951. *Fifty Pame Imperatives*. Ms. SIL, Language & Culture Archives, https://www.sil.org/system/files/reapdata/15/83/01/15830138421724684897123542016095536365/Fift y_Pame_Imperatives.pdf (26.11.2022).

---

[1] ATAM = ASPECTUALITY, TEMPORALITY, ACTIONALITY, MODALITY.

Pragmatische Definitheit in zwei samojedischen Sprachen: eine vergleichende Fallstudie zu Selkupisch und Nenzisch

Josefina Budzisch (Universität Hamburg)

Unter Anwendung der Taxonomie von Hawkins (1978: 106ff.) besteht das Ziel der Untersuchung darin, einen Überblick über die Darstellung anaphorischer, direkt situationsbezogener und assoziativer anaphorischer Verweise im Selkupischen und Nenzischen (beide < samojedisch < uralisch) zu präsentieren und die verschiedenen in Gebrauch befindlichen Strategien zu untersuchen und diese miteinander zu vergleichen.

Diese Studie ist eine korpusbasierte Analyse. Die hier verwendeten Selkupischkorpora Budzisch et al. (2019) und Brykina et al. (2020) bilden sowohl das Nord-, Zentral- als auch das Südselkupische an und umfassen insgesamt 409 Texte (17.092 Sätze, 97.562 Tokens). Das Nenzischkorpus (Budzisch – Wagner-Nagy 2024) entsteht derzeit in Hamburg und enthält Texte aus sowohl dem Tundra- als auch dem Waldnenzischen (157 Texte, 14.704 Sätze, 89.452 Tokens).

Untersuchungen zum Selkupisch (Budzisch 2021) haben gezeigt, dass die Bandbreite der Möglichkeiten vor allem für anaphorische Verweise relativ groß ist. So werden neben Personal- und Demonstrativpronomina (1) auch das Possessivsuffix dritte Singular (2) genutzt – diese aber in geringer Häufigkeit. Zudem finden sich im Selkupischen auch eine große Zahl bereits eingeführter Referenten, die nicht diesbezüglich markiert sind (3).

(1) ***Na***   *šoːqor-t*   *paːr-o-nd*   *siga-l-ba-dɨt.*
     DEM   stove-GEN   top-EP-ILL   climb-INCH-PST.REP-3PL
     'They climbed on top of the oven [which is been introduced before].'
     (Budzisch et al. (2019); TMR_1981_Robbers_flk.044)

(2) *Ašša*   *kuntɨ*   *ɛː-ŋa*   *tama-iː-**tɨ***   *muntɨk*   *ürɨ-jɔː-tɨt.*
     NEG   long   be-CO.3SG.S   mouse-PL.POSS-**POSS.3SG**   all   leave-CO-3PL
     '[The czar's daughter saw: There came so many mice, that they surrounded Itja.] Soon all these mice disappear.'
     (Budzisch et al. (2019); BEP_1977_Itja3_flk.030)

(3) *Paja*   *tüː-p*   *čadɨ-gu*   *laqqa-tɨ-mba.*
     old.woman   fire-ACC   light-INF   begin-IPFV-PST.REP.3SG.S
     'The old woman [present in the discourse beforehand] began to light a fire.'
     (Budzisch et al. (2019); ChDN_1983_MistressOfFire_flk.056)

Das Nenzische ist bezüglich pragmatischer Definitheit nicht umfassend erforscht. Zu Einzelaspekten des Nenzischen, wie dem Gebrauch von Possessivsuffixen finden sich bereits Untersuchungen (Körtvély 2010, Nikolaeva 2014), wobei in den vorliegenden Studien ausschließlich Tundranenzisch als Referenz genutzt wird, das Waldnenzische ist in dieser Hinsicht als unerforscht anzusehen. Eine erste Korpusauswertung zeigt aber, dass sowohl Tundra- als auch Waldnenzisch anaphorische Verweise ebenfalls mit Personal- und Demonstrativpronomina (4) verwenden, aber eben auch Possessivsuffixe (5) – und dies in deutlich größerer Zahl als im Selkupischen nachgewiesen werden konnte; zudem kann in diesen Sprachen nicht nur POSS.3SG sondern auch POSS.2SG genutzt werden. Zudem legt eine Pilotstudie nahe, dass es im Nenzischen im Vergleich zum Selkupischen weniger pragmatisch definite Referenten gibt, die nicht diesbezüglich markiert sind. Dies bedarf noch weitere Forschung.

(4) *Tajn'aʔ*   ***čiki-ʔ***   *n'eša-ʔ*   *kä-ʔm'a-xat-ʔe*   […]
     then   **DEM**-NOM.PL   person-NOM.PL   go.away-ACT-ABL.SG-EXCL
     'Then, after these persons had left […]'

(Budzisch – Wagner-Nagy (2024); AAK_200311_MyLife_nar-150)
(5) *Kad'a,*   *n'e-ł*   *ŋeta-ŋa-ta,*   *ŋeta-ŋa-ta.*
go.3SG.S   woman-**NOM.SG.2SG**   wait-CO-3SG.SG.O   wait-CO-3SG.SG.O
'He went away, but his wife [introduced before] is still waiting for him.'
(Budzisch – Wagner-Nagy (2024); TPG_2002_PakFamily_flk)

Es zeigt sich also einerseits das Selkupisch sich atypisch zu bisherigen Beschreibungen verhält, Nenzisch aber nicht (Nikolaeva 2003) und das sowohl Tundra- als auch Waldnenzisch in das für nordsamojedisch beschriebene Muster (Siegl 2015) passen, sodass hier eine Trennung beobachtet werden kann, die sich auch in der Darstellung von direkt situationsbezogenen und assoziativer anaphorischen Verweisen vermuten lässt, was eine Fragestellung ist, die die vorliegende Untersuchung beantworten soll.

Literatur

Brykina, Maria – Svetlana Orlova – Beáta Wagner-Nagy 2020. INEL Selkup Corpus. Version 1.0. Publication date 2020-06-30. Archived in Hamburger Zentrum für Sprachkorpora. http://hdl.handle.net/11022/0000-0007-E1D5-A. In Wagner-Nagy, Beáta – Alexandre Arkhipov – Anne Ferger – Daniel Jettka – Timm Lehmberg (eds.): *The INEL corpora of indigenous Northern Eurasian languages.*

Budzisch, Josefina – Anja Harder – Beáta Wagner-Nagy 2019. *Selkup Language Corpus* (SLC). Archived in Hamburger Zentrum für Sprachkorpora. Version 1.0.0. Publication date 2019-02-08. http://hdl.handle.net/11022/0000-0007-D009-4 .

Budzisch, Josefina 2021. Definitheit im Selkupischen. (*Studia Uralo-altaica* 55). Szeged: University of Szeged, Department of Altaic Studies and Department of Finno-Ugrian Philology.

Budzisch, Josefina – Beáta Wagner Nagy 2024. *Nenets Corpus.* (INEL-Corpus, in Vorbereitung)

Hawkins, John 1978. *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction.* London: Croom Helm.

Körtvély, Erika 2010. On the Function of Possessive Suffixes in Tundra Nenets: Possession and Semantic-Pragmatic Definiteness. *Finnisch-Ugrische Mitteilungen* 32/33, 321–343.

Nikolaeva, Irina 2003. Possessive affixes as markers of information structuring: Evidence from Uralic. In Suihkonen, Pirkko – Bernd Comrie (eds.): *International Symposium on Deictic Systems and Quantification in Languages spoken in Europe and North and Central Asia.* Iževsk – Leipzig: Udmurt State University – Max Planck Institute of Evolutionary Anthropology, 130–145.

Nikolaeva, Irina 2014. *A grammar of Tundra Nenets* (Mouton Grammar Library 65). Berlin: de Gruyter.

Siegl, Florian 2015. The structure of noun phrases with referential px.2p in Northern Samoyedic. *Tomsk Journal of Linguistics and Anthropology* 1, 21–31.

Die Fokuspartikel *di* im Nuristani Kalasha

Pascal Coenen

Universität zu Köln

Der vorliegende Beitrag beschäftigt sich mit der Fokuspartikel *di* im Nuristani Kalasha (auch Waigali genannt), insbesondere im Dialekt von Nisheygram. Nuristani Kalasha gehört zu den Nuristani-Sprachen, einer Untergruppe des indoiranischen Zweigs der indogermanischen Sprachfamilie. Eine Beschreibung der Funktionen dieser Partikel findet sich in der Grammatik von Degener (1998), die auch eine Sammlung der dokumentierten Texte des Dialekts von Nisheygram sowie ihrer deutschen Übersetzung enthält. Gemäß Degener (1998: 161; 169f.) fungiert *di* als additive Fokuspartikel, die, ähnlich wie das deutsche *auch*, in skalaren sowie in nicht skalaren Kontexten vorkommen kann; des Weiteren fungiert sie als Modalpartikel mit einer Ähnlichen Funktion wie das deutsche *übrigens*, als koordinierende Konjunktion 'und' (bzw. 'sowohl … als auch', wenn doppelt gesetzt) und als subordinierende Konjunktion 'sowie, sobald', wenn sie mit dem Infinitiv steht.

Im vorliegenden Beitrag wird auf der Basis einer erneuten Analyse aller belege im Dialekt von Nisheygram dafür argumentiert, dass es sich bei *di* um eine polyfunktionale Fokuspartikel, nicht aber um eine Modalpartikel handelt. An einigen Stellen, an denen Degener *di* mit 'übrigens' übersetz, liegt tatsächlich ein Kontrast zwischen dem Wort, auf das *di* folgt, und einem Wort in einem anderen Satz vor, sodass *di* diesen Kontrast zu markieren scheint. Dies ist mit der Beobachtung vereinbar, dass additive Fokuspartikeln in vielen Sprachen auch verwendet werden, um kontrastive Topics zu markieren (Forker 2016: 74–77). An anderen Stellen ist eine additive Interpretation möglich.

Eine weitere Funktion, die bisher nicht erkannt wurde, ist die Markierung von „emphatic assertion of identity" (König 1991: 125–131), d.h. die Partikel zeigt an, dass der Referent des Wortes, das ihr vorausgeht, identisch mit einem anderen Referenten ist. Diese Funktion steht in Verbindung mit dem Gebrauch der Partikel nach Infinitiven. In diesen Fällen interpretiert Degener (1998: 161) *di* als subordinierende Konjunktion mit der Bedeutung 'sowie'. Naheliegender ist aber, dass *di* in diesen Fällen als Fokuspartikel fungiert, die den präzisen Moment des Anfangs der durch den Infinitiv ausgedrückten Verbalhandlung hervorhebt.

Dass *di* für die Koordination von Konstituenten innerhalb eines Satzes verwendet wird, ist aus typlogischer Sicht nicht ungewöhnlich, da es u.a. als additive Fokuspartikel fungiert. Diese werden häufig für emphatische, seltener für einfache Koordination verwendet (Forker 2016: 82–84). Der vorliegende Beitrag argumentiert dafür, dass *di* für einfache Koordination verwendet wird, da *di* auch in Kontexten vorkommt, die klar eine kollektive Lesart fordern, während emphatische Koordination nur eine distributive Interpretation zulässt (Bîlbîie 2008).

Ebenfalls typisch für additive Partikeln ist die Tatsache, dass *di* zur Markierung von konzessiven Konditionalsätzen und bei der Bildung von Indefinitpronomina verwendet wird (Degener 1998: 10f.; Forker 2016), die im vorliegenden Beitrag jeweils genauer betrachtet werden.

Literatur

Bîlbîie, Gabriela (2008). „A Syntactic Account of Romanian Correlative Coordination from a Romance Perspective". In: *Proceedings of the 15th International Conference on Head Driven Phrase Structure Grammar, National Institute of Information and Communications Technology, Keihanna*, Stefan Müller (Hg.), 25–45. Stanford, CA: CSLI Publications.

Degener, Almuth (1998). *Die Sprache von Nisheygram im afghanischen Hindukusch*. Wiesbaden: Harrassowitz Verlag.

Forker, Diana (2016). „Toward a Typology for Additive Markers". *Lingua* 180: 69–100.

König, Ekkehard (1991). *The Meaning of Focus Particles. A Comparative Perspective*. London, New York: Routledge.

Negation in locative and existential clauses:

Information structure and pragmatic constraints

It has long been noted (Givón 1978, Tottie 1991, Miestamo 2005, among others) that negative utterances are only felicitous in specific discourse contexts, which all implicitly or explicitly presuppose the affirmative counterpart of the utterance in question. When looking at locative and existential clauses, this seems to hold only for locative clauses, whereas existential clauses behave differently. Comparing *the books are not on the table* to *there are no books on the table*, the former sentence implies that there are books lying somewhere, which were expected to lie on the table, whereas this need not be the case in the latter sentence. Additionally, the locative clause *the book is not on the table* is pragmatically adequate only if the context somehow suggests that the book principally might be on the table (Miestamo 2005: 197–198).

The aim of this talk is to investigate the information-structural properties of negative locative and existential clauses. I will show that the typical argument-focus pattern of locative clauses constraints their possibilities of negation more than the typical sentence-focus pattern of existential clauses does. As a case in point, the Forest Enets (< Uralic) example (1) describes a situation, in which the speaker together with a friend wanted to go fishing, came to the shore and the speaker's boat was not there. Thus, the situation presupposes that the boat should have been there, and the information that the boat is not at its place is a pragmatically adequate assertion. Furthermore, it evokes the reading that the boat is somewhere else now.

(1) ɔdo-jʔ   naʔa-xane-da   d'agu.
    boat-POSS1SG   place-LOC-POSS3SG   NEG.EX
    'The boat is not in its place.'
    (Forest Enets; INEL Enets Corpus)

In example (2), in contrast, there is no expectation of whether there were berries or not in the given year. Still, the negative existential clause is a perfectly adequate assertion about the given year and the availability of berries at that time.

(2) ɛke   pɔ-xon,   ɛke   pɔ-xon   ɔzi-ʔ   d'agu-č.
    this year-LOC this year-LOC berry-PL   NEG.EX-3PL.PST
    'This year, there were no berries in this year.'
    (Forest Enets; INEL Enets Corpus)

To achieve empirically sound conclusions, the analysis relies on a database containing ca. 7,500 locative, existential and possessive clauses from 16 partly unrelated Siberian languages. In the case of existential clauses, 1198 out of 4042 (29.6%) are negative, whereas in the case of locative clauses, it is only 101 out of 1120 (8.3%), the correlation being statistically significant (Fisher-Freeman-Halton exact: p<0.001).

As I will show in the talk, the information-structural properties of locative and existential clauses and the pragmatic constraints for their negation sketched above can explain the underrepresentation of negative locative clauses in the sample, the main reason being that situations enabling utterances like (1) are apparently more frequent in natural speech than situations enabling utterances like (2).

References:

INEL Enets Corpus = Shluinsky, Andrey & Wagner-Nagy, Beáta. In preparation. *INEL Enets Corpus*.

Givón, Talmy. 1978. Negation in Language: Pragmatics, Function, Ontology. In Cole, Peter (ed.), *Pragmatics*, 69–112. New York: Academic Press.

Miestamo, Matti. 2005. Standard negation: The negation of declarative verbal main clauses in a typological perspective (Empirical Approaches to Language Typology 31). Berlin: Mouton de Gruyter.

Tottie, Gunnel. 1991. Negation in English Speech and Writing: A Study in Variation (Quantitative Analyses of Linguistic Structure 4). San Diego [CA]: Academic Press.

The inclusory construction in Komnzo

Christian Döhler

Berlin-Brandenburg Academy of Sciences and Humanities

This first part of the paper describes the inclusory construction (henceforth IC) in Komnzo, which is unique among the world's languages in assigning a distinct set of pronouns as well as a case marker to this construction type. The second part of paper shows the difference between the IC and other types of coordination (eg conjunctive coordination, comitative markers) through a qualitative corpus study.

ICs can be described as constructions in which "some elements of a larger group are referred to along with the larger group itself" (Singer 2001: 1). This can be seen in the Russian example below in (1), where we have the larger group (henceforth: superset) expressed by the pronoun *my* and the included elements of the larger group (henceforth: subset) expressed by the pronoun *toboj*. Superset and subset expression are connected with the comitative marker *s*.

(1)   Russian (Indo-European; Balto-Slavic)

    a. *my*   *s*   *toboj*
          1PL   with   2SG
          'you and I' (lit. 'we with you')

There have been a number of classifications of ICs in the literature, which focus on different coding dimensions. Schwartz (1988, 1988) draws a distinction between (plural) pronoun-coded versus verb-coded ICs. Lichtenberk (2000) highlights the difference between phrasal versus non-phrasal ICs in one dimension and overt versus non-overt coding of ICs in a second dimension. The Komnzo IC is shown below in (2), where the superset is expressed by the verb inflection (3DU). The Komnzo IC allows for the expression of multiple subsets. One of the subset sets is flagged with the appropriate (core) case, which is always non-singular; the ergative =*é* in (2). The other subset is the included subset, which is flagged with the inclusory case marker =*r*.

(2)   Komnzo (Yam; Tonda)

    a. *maureen=é*   *bi*   *y|näbü/nth*   *kowi=r*
          PN=ERG.NSG   sago(ABS)   3DU>3SG.M/beat   PN=IC.DU
          'Maureen beats the sago with Kowi.' (lit. 'Maureen, they beat the sago, with Kowi.') (Döhler 2018: 277)

In typological work on associative plurals, Moravsik has observed that "no language differentiates inclusory and non-inclusory pronouns by assigning distinct phonological forms to the two" (2003: 485). Khachaturyan (2019) has shown that Southern Mande languages do have a set of distinct inclusory pronouns. Komnzo is then the second attested case, which has a distinct set of inclusory pronouns for all person values and for indefinite, interrogative, and placeholder pronominals (cf. Table 1). Moreover, Komnzo is possibly the only attested case of a language with a distinct case marker for the IC.

Table 1: Inclusory pronominals and inclusory case

		DUAL	PLURAL
PERSONAL PRONOUNS	1	*ninrr*	*ninä*
	2	*bnrr*	*bnä*
	3	*nafrr*	*nafä*
PLACEHOLDER		*bafrr*	*bafä*
INDEFINITE		*nä bunrr*	*nä bunä*
INTERROGATIVE		*mafrr*	*mafä*
CASE ENCLITIC		*=r*	*=ä*

References

Döhler, Christian (2018). *A grammar of Komnzo*. Berlin: Language Science Press. DOI: 10.5281/zenodo.1477799.

Khachaturyan, Maria (2019). "Inclusory pronouns in Mande: The emergence of a typological rarum". In: *Folia Linguistica* 53.1, pp. 87–123.

Lichtenberk, Frantisek (2000). "Inclusory pronominals". In: *Oceanic Linguistics* 39.1, pp. 1–32.

Moravcsik, Edith (2003). "A semantic analysis of associative plurals". In: *Studies in Language* 27.3, pp. 469–503.

Schwartz, Linda (1988a). "Asymmetric feature distribution in pronominal 'coordinations'". In: *Agreement in natural language: Approaches, Theories, Descriptions*. Ed. by Michael Barlow and Charles A. Ferguson. Stanford, CA: Center for the Study of Language and Information Palo Alto, pp. 237–50.

Schwartz, Linda (1988b). "Conditions for verb-coded coordinations". In: *Studies in syntactic typology* 17, pp. 53–73.

Singer, Ruth (2001). "A brief investigation of the inclusory construction in Australian languages". MA thesis. The University of Melbourne.

Variation in gender agreement in South-Iranian Arabic

Dina El Zarka, Sandra Ziagos
Universität Graz

In recent times, the study of agreement in spoken Arabic has experienced an enormous upsurge of interest. A fair number of corpus studies have investigated the intricacies of number agreement and its interaction with gender from a typological perspective (cf. Bettega & d'Anna 2022).

While agreement with plural controllers shows complex patterns of variation across varieties, gender agreement with singular controllers appears to be more straightforward. While in Classical Arabic, gender agreement was optional in verb-initial sentences, in modern spoken varieties, singular nouns that are overtly marked for gender as well as nouns denoting females, paired body parts or some unique referents and place names generally trigger feminine singular gender agreement on all types of targets (cf. Bettega & d'Anna 2022).

Gender agreement with singular controllers has predominantly been investigated for the NP domain, where it has been noted that adjectives with the Nisbasuffix -i do not show a feminine target form in some varieties, specifically in Egyptian Arabic (Mitchell 1973). Mismatches in (gender) agreement in spoken Arabic have been related to lack of individuation (Brustad 2000) and specificity (Hoyt 2000).

Our study investigates agreement with singular feminine nominal controllers in Arabic varieties spoken in South Iran, which are in close contact with Persian, a language lacking grammatical gender. Most speakers are bilingual, fluent in Arabic and Persian. However, for younger speakers and speakers in mostly Persian-speaking environments, Persian is dominant. Our goal was to examine under which circumstances gender agreement is not applied by speakers in discourse.

We annotated 1086 targets of 616 controllers in 14 socio-linguistic interviews from 9 villages for several structural and semantico-pragmatic factors known to affect agreement in addition to sociolinguistic variables (speaker age, gender, education and the density of Arab population). The results of the statsitical analysis showed that all sociolinguistic factors as well as most structural and semantico-pragmatic factors, specifically animacy/humanness, had a significant impact on agreement, corroborating the effect of 'individuation' as suggested by prior authors (cf. Bettega & Leitner, 2019 for plural controllers in Khuzestani Arabic). The results suggest that if there is ongoing gender loss, it is in a very early stage. Nevertheless, the development of a default target gender for non-human controllers may be under way. Furthermore, the frequent occurrence of the sg. m. form *tʃaːn* 'to be' in many different functions (tense marker, existential, conditional marker) may spur the ongoing change.

The following examples illustrate variable agreement within and across speakers.

(1) (afb-bka-ci-si-002_235)

jaʕni	*ħʊrma*	*tɪ-ddaxxalˠ*	*tʃaːn*	*mʊsˤiːbɛ*	*tʃaːn*
DM	woman	3SG.F-interfere.IPFV	be.3SG.M.PFV	calamity[SG.F]	be.3SG.M.PFV

If a woman interfered [with men's business], it was a disaster.

(2) (afb-bka-ci-si-002_333-338)

ʊmm-i…	*t-giʃʃ*	*tʃaːn-at*	*ɪl=ɪħwaːj*

Mother-1SG   3SG.F-sweep.IPFV   be.PFV-3SG.F   DEF=house.PL
My mother … used to sweep the house.

   (3) (afb-bak-ci-si-001_114)

bass   *ɪl=ʊwwɪl*   *dɪnja*   *ħɪlu*   *tʃaːn*
but   first   world.SG.F   beautiful.SG.M   be.PFV.3SG.M
But in the past, life was beautiful.

   (4) (afb-bja-ci-si_001_100)

w=ʊmm-i   *maː*   *xallaː-ni*   *gaːm*   *jɪ-btʃi*
and=mother-1SG   NEG   let.PFV.3SG.M-1SG   get_up.PFV-3SG.M   3SG.M-cry.IPFV
And my mother wouldn't let me, she started crying.

References

Bettega, S. & L. d'Anna. 2022. *Gender and Number Agreement in Arabic*. Leiden: Brill.
Bettega, S. & B. Leitner. 2019. Agreement Patterns in Khuzestani Arabic. Wiener Zeitschrift für die
        Kunde des Morgenlandes 109, 9-37.
Brustad, K. 2000. *The Syntax of Spoken Arabic*. Washington, D.C.: Georgetown University Press.
Comrie, B. 1975. Polite plurals and predicate agreement. *Language* 51, 406-418.
Corbett, G. 1991. *Gender*. Cambridge University Press.
Corbett, G. 2006. *Agreement*. Cambridge University Press.
Hoyt, Frederick. 2002. Impersonal Agreement as a Specificity Effect in Rural Palestinian Arabic, in
        Perspectives on Arabic Linguistics xiii‑xiv, 111‑141. Amsterdam, Philadelphia: Benjamins.
Mitchell, T. F. 1973. Aspects of Concord Revisited, with Special Reference to Sindhi and Cairene
        Arabic. *Archivum Linguisticum* iv, 27‑50.

This paper discusses a larger discourse unit emerging from the interaction of clause chains, tail head linkage and prosodic phrasing in narrative speech of Yali (TNG).

Clause chains are a morphosyntactic construction formed by a sequence of so-called *medial clauses* ended by one *final clause,* where medial clauses use not-fully finite and possibly switch reference-marked verb forms, while final clauses are fully independent. Tail head linkage is a more pragmatic discourse-structuring practice, where (the predicate of) the last clause of one chunk of discourse (the tail) is repeated uninformatively as the first clause of the next chunk (the head). Both of these constructions are very widespread in Papuan languages (see e.g. Fedden 2020, Foley 2000: 383ff on clause chains, de Vries 2005, 2006 on tail head linkage).

Prototypically, clause chains are coupled with a prosodic grouping in Yali, such that all medial clauses or chain-internal intonation units are realised with a final rise or high pitch, while final clauses or chain-ending intonation units are produced with final falling intonation. This prosodic grouping could be called a prosodic finality unit, corresponding to what has been discussed by Gussenhoven as continuation vs. finality intonation (e.g. Gussenhoven 2010). Additionally, final clauses are prototypically also the tail of a tail head linkage construction, whose head is then at the same time the first medial clause of the next chain.

Typically, the three discourse-structuring devices thus converge to segment Yali narratives into paragraph- or episode-like chunks. However, in rarer cases, they may also disalign.

This paper discusses what happens in these cases of misalignment. Do the three constructions still mark discourse boundaries in these cases? If so, in how far are these boundaries different from the paragraph boundaries mentioned above, i.e. from boundaries where all three constructions converge?

References:

De Vries, Lourens. 2005. Towards a Typology of Tail-head linkage in Papuan languages. *Studies in Language* 29(2). 363-384.

De Vries, Lourens. 2006. Areal pragmatics of New Guinea: Thematization, distribution and recapitulative linkage in Papuan narratives. *Journal of Pragmatics* 38. 811-828.

Fedden, Sebastian. 2020. Morphology of Trans New Guinea Languages. In *Oxford Research Encyclopedia of Linguistics*.

Foley, William A. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29. 357-404.

Gussenhoven, Carlos. 2010. The phonology of tone and intonation. Cambridge: Cambridge University Press.

Periphrastic possessive and existential constructions in Oromo

In Oromo (a Cushitic language in Ethiopia), existence or location is expressed by the verb *dʒir-* 'exist/there be' and possession by a transitive verb *k'ab-* 'have'. The former is also employed in some context-specific possessive predication, and the latter (i.e., the possessive verb) has a homophonous referential verb *k'ab-* 'hold, grab' which in turn grammaticalizes to possession (Wakweya, Desalegn and Meyer 2019). Thus, Oromo shows both Action and Location Schemas in light of Heine's (1997) schematic types of possessive constructions.

(1)   *boonsaa-n*   *k'arʃii*   ***k'ab**-at-ee-r-a*
          Bonsa-NOM   money.ABS   hold-MID-PFV.CVB-AUX-3SG.M.IPFV
          'Bonsa has money.'

(2)   *boonsaa*   *bira*   *k'arʃii-n*   ***dʒir**-a*
          Bonsa.ABS   near   money-NOM   exist-3SG.M.IPFV
          'Bonsa has money.'

The two constructions in (1) and (2) are the possessive predications expressing ownership relations, and they remain at odds with kinship or body part relations. There are also parallel constructions of both types in Amharic (the most predominant language in Ethiosemitic).

The action schema is almost non-existent in the whole Ethiosemitic except for the occurrence of the verb *yazə* 'hold, seize' as possessive predicator in Amharic (Baye 1997). This special possessive construction in Amharic as in (3) below resembles the Oromo type given in (1), and they usually involve the perfect aspect marking auxiliaries.

(3)   *kassa bɨzu gənzəb*   ***yɨz**-o-all*
          Kassa much money   hold.3SG.M-CVB-AUX.IPFV
          'Kasa has a lot of money.'

The Action Schema is a Cushitic feature (cf. Thomason 1983) whereas the Location Schema is originally the Ethiosemitic type (Weninger 2011; Crass and Meyer 2008). Hence, the Location Schema in Oromo and the Action Schema in Amharic are parallel developments probably as an instance of the mutual feature exchange between the two languages.

This paper aims to describe syntactic and semantic properties of the Action and Location Schemas in Oromo in relation to their parallel constructions in Amharic as a possible contact phenomenon. This is based on the following main observations:

Besides the standard predicative possession construction in Oromo, there are special patterns involving Action and Location Schemas as in (4) below.

(4)   a.   ***k'ab**-+*AUX   'has held'   > 'have'
          b.   *bira+**dʒir**-*   'exists near'   > 'have'
          c.   *harka+**dʒir**-*   'exists at hand'> 'have'

These non-standard predicative possession constructions in Oromo have similar patterns of grammaticalization with the Amharic types.

These constructions predominantly express ownership relations most often in the temporary and physical possession types. Like the Amharic verb *yazə* 'hold, seize', the Oromo referential verb *k'ab-* 'hold, grab' grammaticalizes to possessive construction in combination with the perfect aspect auxiliary. In the existential possessive constructions, the locative

possessor often takes sentence initial position being followed by an indefinite possessum subject in Oromo, and the same is true for Amharic. Thus, the cross-linguistic effect of word order and definiteness in the predicative possession constructions reported by the linguists like Clark (1978) holds true for both languages. These two predominant languages of Ethiopia are probably having shared developments due to their contact for long centuries.

References

Baye Y. (1997). The verb to have in Amharic. In K. Fukui, E. Kurimoto, and M. Shigeta (Eds.), *Ethiopia in Broader Perspectives* (pp. 619–636). Kyoto: Linguistic und Philology.

Clark, E. V. (1978). Locationals: Existential, Locational and Possessive Constructions. In J. H. Greenberg (Ed.), *Universals of Language, vol. 4, Syntax* (pp. 85–126). Stanford University Press.

Crass, J. and R. Meyer. (2008). Ethiopia. In B. Heine (Ed.), *A Linguistic Geography of Africa* (pp. 228–250). Cambridge: Cambridge University Press.

Heine, B. (1997). *Possession: Cognitive Sources, Forces, and Grammaticalization*. Cambridge: Cambridge University Press.

Thomason, S. G. (1983). Genetic Relationship and the Case of Ma'a (Mbugu). *Studies in African Linguistics*, 14(2), 195–231.

Wakweya, O. Gobena, Desalegn, H. and R. Meyer, (2019). The verb *k'ab-* in Oromo. *Zena Lissan*, 28(1), 102–131.

Weninger, S. (2011). Ethio-Semitic in General. In S. Weninger (Ed.), *The Semitic Languages: An International Handbook* (pp. 1114–1123). Berlin: de Gruyter Mouton.

The four-way meaning of tripartite number: implications for a typology of number morphology

Tom Güldemann and Jan Junglas (Humboldt-Universität zu Berlin)

Since the pioneering article by Dimmendaal (2000), the concept of so-called "tripartite" number marking is frequently used to describe the number systems of a variety of languages in northeastern Africa. The essential components of such systems are illustrated in Table 1 by means of the Kadu language Krongo.

Encoding type	Lexeme	Singulative	Unmarked base	Plural
Unmarked base/**Plural**	'house'		*còorì*	*nóo-còorì*
Singulative/Unmarked base	'mosquito'	*tì.n-kíìŋ*	*kíìŋ*	
Singulative/Plural = **Replacement**	'lion'	*tì-kàamù*		*à-kàamù*

Table 1: Tripartite number marking in Krongo (Reh 1985)

In our talk, we will show that the term "tripartite" for a system as in Table 1 has multiple interpretations. In order to make this more transparent, we restate Table 1 in a more abstract form in Table 2.

Noun type	Encoding	Base + X	Unmarked base(s)	Base + X
Trans-numeral	Ø		Variable number	
	P pattern"			Plurative tantum
	S pattern"	Singulative tantum		
	Suppletion		Singular vs. plural	
Bipartite	P pattern"		Singular	Plurative
	R pattern"	Singulative		Plurative
	S pattern"	Singulative	Plural/collective	
Tripartite	T pattern	Singulative	General number	Plurative

Table 2: The four meanings potentially associated with tripartite number

Dimmendaal's meaning of "tripartite" referred to the existence of three distinct encoding types represented in the three lines within the "Bipartite" domain of Table 2. One can also perceive of tripartiteness systemically concerning the formal paradigmatic system, namely the opposition of the three form types of unmarked vs. singulative vs. plurative, as seen by the three right columns of Table 2 (cf. the frame in the first line). At least in some relevant languages (e.g., in the Cushitic family), tripartiteness can also be identified lexically, when

individual lexical items have all three distinct forms of the system (cf. the last line of Table 2). Finally, tripartiteness can pertain to the lexicon in that nouns can show three different ways of morphological behavior regarding the number feature (cf. the frame in the first column of Table 2). In discussing the four different meanings implied by "tripartite" number, we try to show the relevance of these interesting number systems for cross-linguistic research and propose to refine number-related terminology in general.

Dimmendaal, Gerrit J. 2000. Number marking and noun categorization in Nilo-Saharan languages. Anthropological Linguistics 42,2: 214-261.

Reh, Mechthild. 1985. Die Krongo-Sprache (Nìinò Mó-dì): Beschreibung, Texte, Wörterverzeichnis. Kölner Beiträge zur Afrikanistik 12. Berlin: Dietrich Reimer.

Morphological complexity is mostly due to element order rather than to complexification

In this provocative talk, I point out that the prevailing view on "morphological complexity" has ignored the crucial role of element order and has thus gone seriously astray. While earlier work such as Greenberg (1960) confined itself to observing differences in "morphological complexity", work of the last two decades has tended to make claims about exteral causal factors. In particular, McWhorter (2003; 2011; 2016), Lupyan & Dale (2010) and Trudgill (2011) have argued that particular sociolinguistic situations favour the development of "morphologically simple" languages, while others tend to lead to "complexification".

The causal mechanism of simplification through adult second langage acquisition seems plausible at first, but this literature almost never asks whether it is specifically "morphology" that is hard for adult learners to acquire or whether grammatical markers more generally present difficulties. So what exactly is "morphology", and "morphological complexity"?

Most straightforwardly, morphology is the structure of words, and a morphologically complex word is one that has many constituent elements or requires many rules to be generated. But how are word boundaries drawn? There seems to be fairly widespread agreement in practice that affixes exhibit host-class specificity (non-promiscuity; Zwicky & Pullum 1983; Haspelmath 2021), so that, e.g., the English definite article *the* is not a prefix (as it can occur not only before nouns, but also adjectives: *the+cup*, *the+new cup*), while the Swedish definite article *-en* is a suffix (as it occurs only after nouns: *kopp+en* 'the cup', *den nya kopp+en* 'the new cup', *kopp+en där borta* 'the cup over there'). If this is adopted as the key criterion, then it is clear that the affix status of a grammatical marker is co-determined by element order in a wide range of contexts. For the following pairs of possibilities (where the grammatical marker is boldfaced), the next page contains a range of exemplifying languages.

	AFFIX	**NON-AFFIX**
(A)	***flag*** + NOUN (+ adjective)	***flag*** (+ adjective) + NOUN
(B)	(adjective +) NOUN + ***flag***	NOUN (+ adjective) + ***flag***
(C)	(adjective +) NOUN + ***plural***	NOUN (+ adjective) + ***plural***
(D)	***aspect*** + VERB (+ adverb)	***aspect*** (+ adverb) + VERB
(E)	(object +) VERB + ***relativizer***	VERB (+ object) + ***relativizer***
(F)	subject + VERB + ***complementizer***	***complementizer*** + subject + VERB

Flags and number markers can be affixes only if adjectives do not intervene, aspect markers can be affixes if adverbs do not intervene, and so on. Thus, affixhood is dependent on element order.

One might try to adopt other tests for distinguishing affixes from non-affixes, or hope to rely on test batteries, and I will address a few such alternatives in the talk. But none of them have been successful, and it is only for root changes that we have (near-)consensus about their morphological status. Thus, unless one wants to drastically restrict "morphological complexity" to root-change (or endophonic) patterns, element order remains a crucial ingredient. The claims of sociolinguistic typology (which no doubt have some prima facie plausibility) thus need to be approached in a more rigorous fashion in the future.

Finally, I point out that sometimes "morphology" seems to be equated with grammatical marking, which would of course change the claims drastically. There could then be no claim of any relation between sociotypes and word structure types, e.g. no relation between sociotypes and "polysynthesis" (e.g. Fortescue et al. 2017).

AFFIX	NON-AFFIX

(A) *flag* + NOUN (+ adjective)   *flag* (+ adjective) + NOUN
    Hebrew **be**+*vatim gdolim*   English **in+** *big houses*

(B) (adjective +) NOUN + ***flag***   NOUN (+ adjective) + ***flag***
    Turkish *büyük evler+**de*** 'in big houses'   Basque *etxe handi **+etan***

(C) (adjective +) NOUN + ***plural***   NOUN (+ adjective) + ***plural***
    Hungarian *nagy fá+**k***   Kimaghama *do mamu **+ragha***
        big tree+PL   tree big PL (Dryer 1989: 883)

(D) ***aspect*** + VERB (+ adverb)   ***aspect*** (+ adverb) + VERB
    Papiá Kristang *eli **ta**+les buku agora*   English *she **is+** now reading a book.*
        she ASP+read book now   (Baxter 1988: 128-129)

(E) (object +) VERB + ***relativizer***   VERB (+ object) + ***relativizer***
    Turkish *pencere-yi aça+**n** kadın*   Mandarin *dǎkāi chuānghù **+de** nǔrén*
        window-ACC open+REL woman   open window rel woman
    'the woman who opened the window'

(F) subject + verb + ***complementizer***   ***complementizer*** + subject + verb
    Basque *etxe-ra doa+**la***   Indonesian ***bahwa**+ dia akan pulang*
        'that she is going home'   that   she PROG go.home

References

Baxter, Alan N. 1988. *A grammar of Kristang (Malacca Creole Portuguese)* (Pacific Linguistics B 95). Canberra: Pacific Linguistics, Australian National University.

Dryer, Matthew. 1989. Plural words. *Linguistics* 27. 865–896. (doi:10.1515/ling.1989.27.5.865)

Fortescue, Michael D & Mithun, Marianne & Evans, Nicholas (eds.). 2017. *The Oxford handbook of polysynthesis*. Oxford: Oxford University Press.

Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3). 178–194.

Haspelmath, Martin. 2021. Bound forms, welded forms, and affixes: Basic concepts for morphological comparison. *Voprosy Jazykoznanija* 2021(1). 7–28. (doi:10.31857/0373-658X.2021.1.7-28)

Lupyan, Gary & Dale, Rick. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5(1). e8559. (doi:10.1371/journal.pone.0008559)

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2–3). 125–166.

McWhorter, John H. 2011. *Linguistic simplicity and complexity: Why do languages undress?* Berlin: De Gruyter Mouton.

McWhorter, John H. 2016. Is radical analyticity normal? Implications of Niger-Congo and Southeast Asia for typology and dichronic theory. In van Gelderen, Elly (ed.), *Cyclical change continued*, 49–92. Amsterdam: Benjamins.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.

Zwicky, Arnold M. & Pullum, Geoffrey K. 1983. Cliticization vs. inflection: English n't. *Language* 59(3). 502–513.

VeLePa: a Verbal Lexicon of Pame

Borja Herce

University of Zurich
borjaherce@gmail.com

Abstract

This paper presents VeLePa, an inflected verbal lexicon of Central Pame (pbs, cent2154), an Otomanguean language from Mexico. This resource contains 12528 words in phonological form representing the complete inflectional paradigms of 216 verbs, supplemented with use frequencies. Computer-operable (CLDF) inflected lexicons of non-WEIRD underresourced languages are urgently needed to expand digital capacities in this languages (e.g. in NLP). VeLePa contributes to this, and does so with data from a language which is morphologically extraordinary, with unusually high levels of irregularity and multiple conjugations at various loci within the word: prefixes, stems, tone, and suffixes constitute different albeit interrelated subsystems of inflection.

Keywords: under-resourced languages, morphology, complexity, inflected lexicon, entropy

1. Introduction

Central Pame is an indigenous Mesoamerican language spoken by around 5000 people in and around Santa María Acapulco (San Luis Potosí, Mexico). The language is still acquired as a first language by children in various communities, but is endangered by the expansion of Spanish. The language lacks a standard written form, and extant documentation (e.g. Gibson & Bartholomew, 1979; Hurch; 2022) is insufficient, usually undigitized, and computationally largely unusable.

The language, however, like others in its family (e.g. Chichimec, see Palancar & Avelino, 2019; Herce, 2022) is a treasure trove of morphological complexity, due to the combination of the following two traits:

- High levels of irregularity, with many small inflection classes, many uniquely-behaving verbs, and suppletion.
- A morphological realization of subject and tense information which is distributed along the word into multiple inflectional layers: prefixes, tone, stem, and suffixes.

These properties make the system highly interesting and challenging to theoretical morphology, as well as to NLP. Adding this language to cross-linguistic databases like Unimorph (see McCarthy et al., 2020) and to morphological reinflection tasks would make these more representative of human language diversity and its limits.

2. Building VeLePa

To build an inflected lexicon of Central Pame verbs the first thing we need is language documentation. Although some inflectional paradigms have been collected by SIL missionaries around 70 years ago (Gibson, 1950), these are insufficient in both number and quality. Over the last four years, I have been documenting the language together with native speakers, through the elicitation of all inflected forms from a large number of verbs.

Every single one of the 12528 inflected forms that VeLePa contains (all 58 forms from 216 verbs) has been independently elicited (i.e. never extrapolated from other forms, as is often the case of these recources) and checked multiple times to avoid mistakes and inconsistencies (e.g. in the treatment of synonymous inflected forms lie *dived~dove*). This is needed, first, because the language demands it. 74% of these words are morphologically different, and syncretisms are never the result of different values being systematically the same across all lemmas as in other languages (e.g. English *do* 1SG.PRS, *do* 2SG.PRS, *do* 1PL.PRS, *do* 2PL.PRS, *do* 3PL.PRS). Secondly, given the large degree of irregularity in the language, one can hardly ever be sure to be able to fill out the complete paradigm correctly from a subset of its cells. Eliciting every single form prevents underestimating complexity. At the same time, however, because VeLePa has been built with computational analysis in mind, cross-speaker and intra-speaker variability and free variation had to be ironed out in a way that this does not lead to an *over*estimation of complexity. Although crucial, this type of quality controls are not always discussed and implemented in the compilation of inflected lexicons, particularly those from indigenous languages, as these tend to be produced by documentary linguists for whom the computational use of these resources is not a priority.

Given the absence of a standard of the language, and the unsuitability[1] of the orthography generally in use in the community, forms are represented in VeLePa in phonological transcription. Tones (High [H], Low [L], and Falling [F]) are indicated immediately after the (lowest) vowel of the syllable where they occur. Consonant gemination is indicated through a doubling of the corresponding consonant. To facilitate analysis, segmentations of prefix and stem have been included (indicated by "-"), as well as zero prefixes (indicated by "0"). These can be deleted if morphological decomposition is not needed. Other transcription choices are IPA-compliant. Typical forms are hence

---

[1] While these are phonemic in the language, neither tone nor vowel nasalization nor consonant gemination are consistently represented in writing.

to-hoHʔo, 0-mbãLnʔ, laH-ppo, la-hõFlʔ, etc. or from a single verb la-nõH , ta-nõHn, ki-ŋõHik, 0-nõH, etc.

Every inflected form is, of course, tagged for its lemma (e.g. 'play') and morphosyntactic values (e.g. 1SG.PRS). As a further feature of interest to computational morphologists, for example those interested in the Paradigm Cell Filling Problem in a naturalistic setting, (see Ackerman et al., 2009; Blevins et al., 2017), I also provide a use frequency estimate of the different lemmas (see Figure 1, frequency estimated in number of tokens per million words) and morphosyntactic values (see Figure 2, frequency estimated as proportion of verbal tokens). These were derived from the frequency of forms in extant Central Pame texts (see Gibson et al, 1963; Gibson, 1966; Hurch 2022), supplemented with subjective frequency estimates from native speakers (see Carrol, 1971) due to the small size of the available corpus (only 1171 verbal tokens) and its unbalanced thematic and genre composition (with an overrepresentation of narrative).
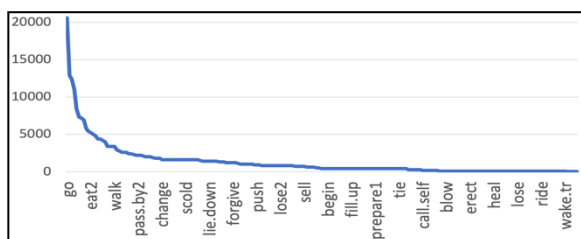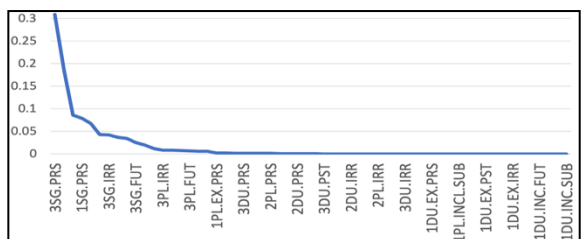

Figure 1: Frequency rankings of lemmas

Figure 2: Frequency rankings of values

3. Analysis of system complexity

On the basis of VeLePa, freely available online, we analyze the morphological complexity and the predictability of the inflectional system as per the Paradigm Cell Filling Problem (see Ackerman et al., 2009). As mentioned in the introduction, one of the key idiosyncratic features of the language is the relative independence of prefixal, suffixal, tonal, and stem morphology. These four layers are analyzed separately below, through the following software:

- Qumín (Beniamine, 2018), for the automatic extraction of morphological alternations, and for the calculation of Information-Theoretic measures (e.g. conditional entropy of one form given another).
- Principal Parts Analyzer (Stump & Finkel, 2013), for the calculation of Set-Theoretic measures like the number of principal parts (i.e. the lowest number of forms required to predict the complete paradigm).

3.1 Prefixes

Despite their exhuberant allomorphy and the presence of stem-initial alternations, prefixes are straightforward to segment from stems. As the exemplary forms in Section 2 suggest, the prefix is the most changeable part of the word, and setting aside cases of zero-prefixed forms, corresponds generally to the first syllable of the word. Given this identification of prefixes, Pame verbs classify into 22 different inflectional classes, with a few comparatively frequent ones (see Table 1), and a long tail of (12) verbs which are prefixally unlike any other in the database.

type freq.	85	51	24	10	9	6	5	5
1SG.PRS	la	to	ti	la	la	ti	la	to
1DU.EX.PRS	ta	to	ti	ta	ta	ti	ta	to
1DU.INC.PR	ta	to	ti	ta	ta	ti	ta	to
1PL.EX.PRS	ta	to	ti	∅	ta	ti	wa	to
1PL.INC.PRS	ta	to	ti	∅	ta	ti	wa	to
2SG.PRS	ki	to	ti	ki	ki	ti	ki	la
2DU.PRS	ki	to	ti	ta	ki	ti	ta	la
2PL.PRS	ki	to	ti	∅	ki	ti	wa	la
3SG.PRS	wa	lo	li	∅	∅	li	∅	wa
3DU.PRS	wa	lo	li	∅	∅	li	∅	wa
3PL.PRS	∅	wa	ti	∅	∅	li	wa	wa

Table 1: Present prefixes of the 8 largest classes

As Table 1 shows,11 values of person-number are distinguished in the language, over 6 values of tense-aspect-mood. Due to the incompatibility of 1st and 3rd persons with the imperative mood, 58 values/cells exist in the Pame verb's paradigm. These fall into 39 areas of mutual interpredictability (see Table 2). These are those areas where the content of one cell (e.g. the 1PL.EX.PRS) allows to to predict that of another (e.g. 1PL.INC.PRS) and *vice versa*. In Pame this tends to mean their forms are always the same (e.g. ta/ta, to/to, ti/ti., ∅/∅, or wa/wa in Table 1).

	PRS	PST	IRR	SUB	FUT	IMP
1SG	1	9	16	24	32	-
1DU.EX	2		17	25		-
1DU.INC						-
1PL.EX	3	10	18	26	33	-
1PL.INC						-
2SG	4	11	19	27	34	37
2DU	5	12	20	28	32	38
2PL	6	13	21	29	33	39
3SG	7	14	22	30	35	-
3DU						-
3PL	8	15	23	31	36	-

Table 2: Prefix interpredictability areas

The average conditional entropy (i.e. a measure of the uncertainty involved in predicting one form from another) is 0.52 bits. On a different metric of complexity, 5 static principal parts are needed to predict the entire paradigm. These speak of the complexity of prefixal inflection in Central Pame, which is, however, lower than that of the other inflectional layers/subsystems in the language that will be presented in the next sections.

3.2 Stems

While all Pame verbs show prefixal and suffixal inflection, only most (96.3%) display stem alternation. Barring cases of suppletion, which occurs in twelve verbs, generally with different roots in SG/DU and PL, most of the morphological action in stems occurs on their consonantal onset. Sometimes, particularly in the 3PL across tenses, it involves the addition of segments, some other times it involves gemination, sometimes segmental changes, etc. These occur with somewhat recurrent distributions in the paradigm (see a summary of the largest classes in Table 3).

type freq.	16	13	6	5	5	5	5	5
1SG.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
1DU.EX.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
1DU.INC.PR	pp	ʔ	ʔu	h	kk	pp	tt	tt
1PL.EX.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
1PL.INC.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
2SG.PRS	ppy	ʔy	ʔu	h	kky	pp	kky	kky
2DU.PRS	ppy	ʔy	ʔu	h	kky	pp	kky	kky
2PL.PRS	ppy	ʔy	ʔu	h	kky	pp	kky	kky
3SG.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
3DU.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
3PL.PRS	b	lʔ	tʔ	th	kh	pp	lh	lʔ
1SG.PST	w	ʔu	ʔu	h	ku	pp	t	t

Table 3: Present stem onsets of the 8 largest classes

Given the regularities in the distribution over values of different alternations, the 58 cells of the Pame verb paradigm are grouped into 29 interpredictability areas (see Table 4). The average conditional entropies between them is 0.63 bits, and 6 principal parts are minimally needed to be able to predict the complete stem paradigm without uncertainty.

	PRS	PST	IRR	SUB	FUT	IMP
1SG						-
1DU.EX	1	7	13		22	
1DU.INC						
1PL.EX	2	8	14	19	23	
1PL.INC						
2SG	3	9	15		24	28
2DU						
2PL	4	10	16	20	25	29
3SG	5	11	17	21	26	-
3DU						
3PL	6	12	18		27	

Table 4: Stem interpredictability areas

3.3 Tones

Tone (high, falling, or low) occurs in Pame in the stressed syllable, which can be either the final one (i.e. the root), or the penultimate (i.e. the prefix). Tone and stress are further intertwined in the language in that only the high tone occurs when the stressed syllable is the penultimate. The result is that only 4 tone-stress profiles are possible in any given word.

While all or most Pame verbs are inflectable in the other morphological layers, tone is different in that most verbs (66.2%) have a single tone across the paradigm (see the 4 largest classes in Table 5). Despite this, the PCFP is a considerable challenge because there is no way to predict, from the tonal value of a given form, whether this same tone will be found across the paradigm or in specific domains only, of which 19 exist (see Table 6).

type freq.	52	47	25	18	8	5	5	4
3SG.PRS	-L	-H	H-	-F	-L	H-	-H	-H
3DU.PRS	-L	-H	H-	-F	-L	H-	-H	-H
3PL.PRS	-L	-H	H-	-F	-L	-H	-H	-H
1SG.PST	-L	-H	H-	-F	-L	-H	-L	-H
1DU.EX.PST	-L	-H	H-	-F	-L	-H	-L	-H
1DU.INC.PST	-L	-H	H-	-F	-L	-H	-L	-H
1PL.EX.PST	-L	-H	H-	-F	-L	-H	-L	-H
1PL.INC.PST	-L	-H	H-	-F	-L	-H	-L	-H
2SG.PST	-L	-H	H-	-F	-F	-H	-F	-L
2DU.PST	-L	-H	H-	-F	-F	-H	-F	-L
2PL.PST	-L	-H	H-	-F	-F	-H	-F	-L

Table 5: Tones of the 8 largest classes

	PRS	PST	IRR	SUB	FUT	IMP
1SG						-
1DU.EX	1	6	10		16	
1DU.INC						
1PL.EX	2	7	11		17	
1PL.INC						
2SG	3	8	12		18	12
2DU						
2PL	4	9	13		19	13
3SG	1	6	14			-
3DU						
3PL	5	7	15			

Table 4: Stem interpredictability areas

Despite the small number of possible values of tone, the average conditional entropy between these domains is 1.01, and one would need minimally 7 principal parts to be able to predict with certainty the tone of every inflected form. These values are the highest among all four inflectional layers.

3.4 Suffixes

While prefixes, stems, and tones encode, often redundantly, different values of subject person-number, and tense-aspect-mood, suffixes tend to encode person-number almost exclusively. Pame suffixes are always non-syllabic, attaching as a syllable coda when the stem finishes in a vowel (e.g. kowwaL +i > kowwaLi; kowwaL +nʔ > kowwaLnʔ) but modifying the stem ending when the root already has a coda (e.g. toŋgoãHn +i > toŋgoãHiŋ, toŋgoãHn +nʔ > toŋgoãHnʔ). This gives rise to unpredictability in that, given a suffixed form (e.g. one which contains an underlying suffix -nʔ), it cannot be known what the unsuffixed form is (e.g. ∅ vs -n in the verbs above).

Alongside this source "superficial" unpredictability, suffixes also change from verb to verb. As the forms in Table 7 show, some have a 2DU suffix -k while others do not, and some have a 3PL suffix -t while others do not. The two sources of unpredictability combine to generate a PCFP challenge comparable to the other inflectional layers, with 14 areas of interpredictability (see Table 8), 0.62 bits of average conditional entropy, and 6 static principal parts.

type freq.	50	27	22	8	8	6	6	6
1SG.PRS	∅	ŋ	?	n	∅	?	ŋ	t
1DU.EX.PRS	m?	m?	m?	n?	m?	m?	m?	n?
1DU.INC.PR	∅	∅	?	ŋ	∅	?	∅	t
1PL.EX.PRS	n?	n?	n?	n?	n?	n?	n?	n?
1PL.INC.PRS	n	n	n	n	n	n	n	n
2SG.PRS	∅	ŋ	?	n	∅	?	ŋ	t
2DU.PRS	∅	∅	?	ŋ	k	?k	∅	t
2PL.PRS	n	n	n?	n	n	n?	n	n
3SG.PRS	∅	ŋ	?	n	∅	?	ŋ	t
3DU.PRS	∅	∅	?	ŋ	∅	?	∅	t
3PL.PRS	∅	ŋ	?	n	t	?	nt	t

Table 7: Suffixes of the 8 largest classes

	PRS	PST	IRR	SUB	FUT	IMP
1SG	1	10	1			-
1DU.EX	2					-
1DU.INC	3	11	3			-
1PL.EX	4					-
1PL.INC	5					-
2SG	6	12	6			
2DU	7		7			
2PL	8					
3SG	1	10	1			-
3DU	3	11	3			-
3PL	9					-

Table 8: Suffixal interpredictability areas

4. Conclusion

An inflectional system with the complexity of any of these layers would be considered quite complex. The (in)famous Latin verbs, for example, have 5 principal parts, 0.28 bits of average conditional entropies, and 15 zones of interpredictability (see Pellegrini, 2020), yet this is almost consistently simpler than any of the inflectional subsystems that coexist within Pame verbs. The overall system, hence, would appear to test the very limits of human linguistic cognition. How do speakers manage to successfully learn and use a system like this? The answer might lie in predictability *between* inflectional layers. While that between cells is more often explored, this does not mean that predictability between different parts of a single word plays no role. Preliminary assessment of how much information one layer provides about another in Pame can be obtained from Normalized Mutual Information (NMU), calculated through the R package `aricode` (Chiquet et al., 2020). Results show NMU oscillates between 0.32 and 0.58, which means the lexical classifications of different layers are highly informative about each other.

Beyond these aspects, another challenging aspect of Pame verb morphology is the unsystematic nature of syncretism. While this is frequent in the language (26% of forms), this does not occur systematically, in that there are no cells in the paradigm that are always syncretic. Prefixal inflection classes, for example, (see the largest ones in Table 1) differ not only in their use of different allomorphs, but also in their differential partition of the semantic space. Because the pattern of contrasts is different in every class of verbs, it must make the Paradigm Cell Finding Problem (see Boyé & Schalchli, 2019) extremely challenging. Predicting the lemma and morphosyntactic value of a form from its morphology must also be complicated because of the reuse of the same markers with different functions (see e.g. ∅ and wa in Table 1). These and other aspects of morphological organization in Pame verbs can now be explored through the resource VeLePa.

5. Bibliographical References

Ackerman, F., Blevins, J. P. and Malouf, R. (2009). Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In J. P. Blevins and J. Blevins (Eds.), Analogy in grammar: Form and acquisition: 54–82. Oxford: Oxford University Press.

Beniamine, S. (2018). Classifications flexionnelles. Étude quantitative des structures de paradigmes. PhD diss., Université Sorbonne Paris Cité-Université Paris Diderot.

Blevins, J. P., Milin, P. and Ramscar, M. (2017). The Zipfian paradigm cell filling problem. In Perspectives on morphological organization: 139-158). Brill.

Boyé, G., and Schalchli, G. (2019). Realistic data and paradigms: The paradigm cell finding problem. Morphology, 29, 199-248.

Carroll, J. B. (1971). Measurement properties of subjective magnitude estimates of word frequency. Journal of Verbal Learning & Verbal Behavior, 10, 722-729.

Chiquet, J., Rigaill, G., Sundqvist, M., Dervieux, V., and Bersani, F. (2020). Package 'aricode'. R package version.

Gibson, L. F. (1950). Verb paradigms in Pame. https://www.sil.org/resources/archives/53023.

Gibson, L. F. (1966). The Man Who Abandoned his Wife. A Pame Story. Tlalocan 5, no. 2: 169-177.

Gibson, L. F., Olson, D. and Olson, A. (1963). Four Pame Texts. Tlalocan 4, no. 2: 125-143.

Herce, B. (2022). Possessive inflection in Chichimec inalienable nouns: The morphological organization of a closed irregular class. Studies in Language 46, 4: 901-933.

Hurch, B. (2022). Pame (central) de Santa María Acapulco, Santa Catarina, San Luis Potosí. In Y. Lastra (Ed.) Archivo de Lenguas Indígenas de México. México DF: El Colegio de México.

McCarthy, A. D., Kirov, C., Grella, M., Nidhi, A., Xia, P., Gorman, K., Vylomova, E. et al. (2020). UniMorph 3.0: Universal Morphology. In Proceedings of The 12th language resources and evaluation conference, pp. 3922-3931. ELRA.

Palancar, E. L. and Avelino, H. (2019). Inflectional complexity and verb classes in Chichimec. Amerindia 41: 323-360.

Pellegrini, M. (2020). Using LatInfLexi for an entropy-based assessment of predictability in Latin inflection. In Proceedings of LT4HALA 2020-1st workshop on language technologies for historical and ancient languages (pp. 37-46).

Stump, G. and Finkel, R. A. (2013). Morphological typology: From word to paradigm. CUP.

Numeral classifiers in Bwamu

Cross-linguistically, there are three types of numeral classifiers that can be referred to as mensural (e.g., *glass*), "groupal" (e.g., *flock*), and sortal. While the former two are widespread, the last type of classifiers is a feature best known from languages in Southeast Asia and some adjacent areas. With some exceptions (e.g., Ikoro 1994), this type of classifier is by and large absent from the languages of Africa (Aikhenvald 2000: 122-124; Gil 2013).

This study presents a typological description of the sortal numeral classifier system in the North Central Gur language Bwamu spoken in Burkina Faso. Bwamu is a previously poorly described language for which only few materials are available (e.g., Manessy 1960; Yé 1981). The variety described in this study is the one from Houndé in the province of Tuy that is the native language of the second author of this study.

Altogether, there are four classifiers that, like the articles, distinguish between number and animacy (Table 1). As opposed to languages like Mandarin, there are no sortal classifiers that refer to specific semantic classes based on shape, such as *zhāng* for flat objects.

Table 1: Bwamu classifiers

	singular	plural
things, plants	dὲ	bó
animals, humans	wὲ	bá

Instead of mensural or "groupal" classifiers, the sortal classifiers are used with compounds in which the latter element corresponds to the classifier in other languages like English.

(1) [gni-ni **gnunɛ**]   **bo**   tĩ
     bird-pl group   clf.pl   three
     'three flocks of birds'

Numerals follow the head noun and obligatorily combine with a preceding classifier: N-[Clf-Num]. Classifiers furthermore combine with plural marking on the noun, which is a cross-linguistically dispreferred pattern (Cathcart et al. 2020). The plural marking is a remnant of a formerly more complex noun class system (Beyer 2012).

(2) (hã-ɓɛ)   [**ba**   tĩ]
     woman-pl   clf.pl   three
     'three (women)'

The head noun can be left out, for example in answers to questions. The presence of the classifier also allows the extraction of the head under topicalization, which is impossible for nouns modified by quantifiers or demonstratives that do not have a classifier.

(3) **Livru-ɓɛ**   ĩ   jɛ   [__   [**bo**   tĩ]].
     book-pl   1sg   buy(.pfv)   clf.pl   three
     'Books, I have bought three.'

In combination with other modifiers, there is N-Dem-Adj-[Clf-Num] order, which is cross-linguistically rare (Dryer 2018: 799).

In sum, Bwamu exhibits several typologically interesting features that deserve a comprehensive description. For a full understanding of the system from both areal and typological perspectives, it will be contrasted with classifier systems in the neighboring Mande language Jula, which has only one classifier *den*, as well as Mandarin Chinese, for which data were elicited from native speakers.

References

Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices.* Oxford: Oxford University Press.

Beyer, Klaus. 2012. Bwamu. In Gudrun Miehe, Brigitte Reineke & Kerstin Winkelmann (eds.), *Noun class systems in Gur languages. 2, North Central Gur languages*, 604-616. Köln: Köppe.

Cathcart, Chundra, Andreas Hölzl, Gerhard Jäger, Paul Widmer & Balthasar Bickel. 2020. Numeral classifiers and number marking in Indo-Iranian: A phylogenetic approach. *Language Dynamics and Change* 2020. 1-53.

Dryer, Matthew S. 2018. On the order of demonstrative, numeral, adjective, and noun. *Language* 94. 798-833.

Gil, David. 2013. Numeral classifiers. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/55

Ikoro, Suanu M. 1994. Numeral classifiers in Kana. *Journal of African Languages and Linguistics* 15. 7-28.

Manessy, Gabriel. 1960. *La morphologie du nom en Bwamu (bobo-oulé). Dialecte de Bondoukuy.* Paris: Université de Paris. Faculté des Lettres et Sciences Humaines. (Doctoral dissertation.)

Yé, Vinou Muntu. 1981. *La morphologie du nom et du verbe en bwamu, dialecte de Bagassi.* Université de Nice. (Doctoral dissertation.)

Can an early 19th-Century Manuscript tell us something about the linguistic situation in Fouta Jallon (Guinea) at the time of its production?

Jannis Kostelnik, University of Hamburg | jannis.kostelnik@uni-hamburg.de

Unlike in some regions of Asia or Europe, where many written languages can be traced back at least several centuries, this time-depth of written language is comparatively rare in Sub-Saharan Africa, which makes it harder to reconstruct previous stages of the multilingual situation apparent in large parts of the continent on the one hand and of single languages on the other hand. Notwithstanding that, there are cases in which linguists could deduce information about earlier stages of African languages by reading both figuratively and literally 'between the lines' of written artefacts: The existence of the Northern Songhay language Emghedesie, now extinct, is only known to African Linguistics thanks to the travel diary of the German explorer Heinrich Barth, who described it in 1850 (Lacroix 1981). Similarly, archival notes left behind by his British contemporary and colleague James Richardson have been exploited as a resource for historical studies of the Tuareg variety of the Libyan oasis of Ghat (Souag & Benkato 2022). Most recently, Souag (2023) has collected information on historical sound changes in Songhay languages by evaluating both historic rock inscriptions from modern-day Mali as well as occurrences and descriptions of these languages in West African manuscripts.

This contribution seeks to explore the possibilities of extracting information about the historical linguistic situation in the Futa Jallon (Guinea) from a manuscript kept at Hamburg State and University Library. *Cod. in Scrin. 227a* is a composite manuscript compiling a selection of prayers, talismans and medical recipes in which the main language is Arabic, mixed with a Manding variety bearing closest resemblance to the varieties known today as Maninka and Kakabe. This manuscript is especially suited to such a study, since, unlike other manuscripts of its kind, it has the advantage that its place of production is known and that its production can be dated to a reasonably narrow period of time. However, the contribution also aims to point out a few caveats in such a study, which are conditioned by the script used in the manuscript, its producer's identity and, not least, its genre.

References:

Lacroix, Pierre Francis. 1981. Emghedesie 'Songhay language of Agades', a travers les documents de Barth. In: Société des Africanistes (ed.): *Itinérances... en pays peul et ailleurs: mélanges réunis par les chercheurs de l'ERA 246 du CNRS à la mémoire de Pierre Francis Lacroix*. Paris: Société des Africanistes, pp. 11-19.

Souag, Lameen. 2023. Manuscripts as witnesses to Sound Changes in Songhay. Paper presented at the Conference *Touching, Seeing, Reading, Hearing: Voyage into Islamic Manuscripts of West Africa*, University of Hamburg, 16-18 November 2023.

Souag, Lameen & Adam Benkato. 2022. The Tuareg Dialect of Ghat in 1850. In: *Lamma: A Journal of Libyan Studies*, 2, pp. 9-84.

Vielfaltslinguistik Graz 2024

Lexical aspectual verb classes and differential lexicalisation

Johanna Mattissen (Köln)

Verb-internal lexical and morphological elements and verb-external elements contribute to the aspectual value of the whole clause.

Lexicalised aspectual values of the verb are described here within an enhanced model of the Breuan actional lexical classes (Breu 1996, 1997; cf. Ebert 1995; Mattissen 2003, to appear). A focus will be put on differential lexicalisation with respect to both actional and valency class, as in Nivkh (cf. Gruzdeva 2012), Cayuga (cf. Sasse 2000) or Kartvelian languages (Mattissen 2001, 2003, to appear; Holisky 1981) and with respect to verb/satellite-framing, as in Greek, Spanish or Japanese (cf. Giannakidou/Merchant 1999; Nishida 1994). In addition, binary morphological aspect, telicity and orientation marking on verbs (as in Romance, Slavic (cf. Lindstedt 1985) and Kartvelian languages, Hungarian, and Tagalog (cf. Latrouite 2011)) are taken into account.

It will be argued that verb-external aspectual elements such as incremental/measuring participants, cognate objects, telos/explicit-endpoint participants and results (as in Finnish (cf. Heinämäki 1984), Hungarian, English, Romanian (cf. Farkas 2011) or Basque) do not recategorise the lexically inherent actional class of a verb, but activate its class-defining boundary-and-phase readings in correlation with inherent transitivity and manner of affectedness, and that manner of affectedness is head-marking, i.e. verb-internal, in aspectually head-marking languages.

The model is designed to be universally applicable and to provide a consistent explanation of the aspectual behaviour of verbs, stacking of aspectual marking, of the different readings of perfective and imperfective forms of verbs, their interaction with the affectedness and quantification of nominal participants as well as with various adverbials.

Selected References

Breu, Walter 1996: "Komponentenmodell der Interaktion von Lexik und Aspect". In: Girke, W. (ed.): *Slavistische Linguistik 1995*. München: Sagner. 37-73

Breu, Walter 1997: "Komplexe aktionale Verbklassen, insbesondere Inchoativa". In: Berger, T. / Raecke, J. (eds.): *Slavistische Linguistik 1997*. München: Sagner. 55-80

Ebert, Karen H. 1995: "Ambiguous Perfect-Progressive forms across languages". In: Bertinetto, Pier Marco / Bianchi, Valentina / Dahl, Östen / Squartini, Mario (eds.): *Temporal Reference, Aspect and Actionality*, Vol. 2: Typological Perspectives. Torino: Rosenberg & Sellier. 185-203

Farkas, Imola-Ágnes 2011: "Resultative Expressions in Romanian". *Ianua. Revista Philologica Romanica* 11: 67-88

Giannakidou, Anastasia / Merchant, Jason 1999: "Why Giannis can't scrub his plate clean: On the absence of resultative secondary predication in Greek". In: Moser, Amalia (ed.): *Greek Linguistics: Proceedings of the 3rd International Conference on Greek Linguistics*. Athens: Ellinika Grámmata. 122-134

Gruzdeva, Ekatarina Ju. 2012: "Aspektual'nye klassy glagolov v nivxskom jazyke". In: Kazanskij, N.N. / Plungjan, V.A. (eds.): *Tipologija aspektual'nyh sistem i kategorij*. Sankt-Peterburg: Nauka. 458-533

Heinämäki, Orvokki 1984: Aspect in Finnish. In de Groot, Caspar & Tommola, Hannu (eds.): *Aspect Bound: A Voyage into the Realm of Germanic, Slavonic and Finno-Ugrian Aspectology.* Dordrecht: Foris. 153-177

Holisky, Dee Ann 1981: *Aspect and Georgian Medial Verbs*. Delmar: Caravan

Latrouite, Anja 2011: *Voice and Case in Tagalog: The coding of prominence and orientation*. Düsseldorf: HHU

Lindstedt, Jouko 1985: *On the Semantics of Tense and Aspect in Bulgarian*. Helsinki: University of Helsinki Press

Mattissen, Johanna 2001: "Tense and Aspect in Laz". In: Ebert, Karen H. / Zúñiga, Fernando (eds.): *Arbeiten des Seminars für Allgemeine Sprachwissenschaft (ASAS) 16: Aktionsart and Aspectotemporality in Non-European Languages*. Zürich: Seminar für Allgemeine Sprachwissenschaft. 15-48

Mattissen, Johanna 2003: "Aktionale Klassen in polypersonalen Aspektsprachen - am Beispiel Lasisch". In: Boeder, Winfried (ed.): *Kaukasische Sprachprobleme*. Oldenburg: Carl-von-Ossietzky-Universität Oldenburg. 247-274

Mattissen, Johanna (to appear): "Lexical actional classes in Romance languages in interaction with aspectually relevant material in the clause". In: *Beihefte zur Zeitschrift für romanische Philologie*. Berlin: DeGruyter.

Nishida Chiyo 1994: "The Spanish reflexive clitic *se* as an aspectual class marker". *Linguistics* 32: 425-458

Sasse, Hans-Jürgen 2000: "Aspektsemantik und Lexikonorganisation im Cayuga". In: Breu, Walter (ed.): *Probleme der Interaktion von Lexik und Aspekt* (ILA). Tübingen: Niemeyer. 193-246

The nature of copulas lies in coding predicativity

This paper reports on a study of word order in nominal predicate constructions and suggests that the nature of copulas lies in coding *predicativity*.

While word order typology has been the primary focus in modern linguistic typology, the word order in nominal predicate constructions has received little attention with a few exceptions (e.g, Dryer [1992]; Curnow [1999]; Gong & Uehara [2024]).

Based on an 80-language sample, this study shows that the order of subject and predicate nominal strongly correlates with the order of subject and verb in languages that employ the zero strategy. Meanwhile, the order of subject, predicate nominal, and copula strongly correlates with the order of subject, object, and verb, respectively, in languages using the copula strategy:

(1)   Kuot (Isolate; Chung & Chung 1996: 7,1)

       a.   *migana*   *irie*

           man   3SG.M

           'He is a man.'

       b.   *parak-oŋ*   *ira-ruaŋ*   *kamin*

           eat-3SG.SBJ   father-1SG   yam

           'My father eats yam.'

(2)   Irabu (Japonic; Shimoji 2008: 122, 151)

       a.   *kari=a*   *sinii=du*   *a-tar*

           3SG=TOP   teacher=FOC   COP-PST

           'He is a teacher.'

       b.   *maju=nu*   *jumunu=u=du*   *tur-tar*

           cat=NOM   mouse=ACC=FOC   catch-PST

           'The cat caught a mouse.'

However, this does not imply that the copula belongs to the verbal category. As is well-known, there are non-verbal copulas in the world's languages. Instead, my results suggest that copulas show the ordering properties of verbs regardless of their type (verbal or non-verbal). One of the main reasons for these results is that diachronic sources for copulas, such as pronouns, demonstratives, topic/focus markers, and prepositions—rather than verbs—typically evolve into copulas in verb-medial languages. Pronouns, demonstratives, and topic/focus markers follow the subject, while prepositions precede the predicate nominal. As a result, in verb-medial languages, copulas derived from these elements develop to correlate with verbs, as in (3).

(3)   Modern Hebrew: Copula derived from a pronoun (Katz 1996: 90)
       'Ata   *hu*   *ha-'iš*
       thou.M   COP   DEF-man
       'You are the man.'

Concerning the question of why copulas are derived from elements other than verbs, typically in verb-medial languages, Hengeveld et al. (2004) explain this by citing ambiguity. However, my study does not support their claim because many Australian languages do not use a copula even though their word order is flexible (Dixon 2002: 78, 240). This implies that ambiguity is not always avoided, as also observed by Piantadosi et al. (2012) and Wasow (2015). Instead, this study suggests that the nature of copulas lies in coding *predicativity*. Languages that do not use copulas can employ nominals as predicates, like verbs. In such languages, a predicate nominal can occupy the position of a verb. On the other hand, languages that use copulas employ nominals as predicates with the assistance of copulas. In such languages, copulas allow nominals to stand as predicates to nouns. This conclusion is consistent with Stassen (1997: 106-112) and Croft (2022: 295-296), as the zero strategy is primarily used for identification, not predication. Also, this conclusion empirically supports the definition of the copula given by Haspelmath (2024).

References

Chung, Kyung-Ja & Chul-Hwa Chung. 1996. Kuot grammar essentials. In John M. Clifton (ed.), *Two non-Austronesian grammars from the islands* (Data Papers on Papua New Guinea Languages), vol. 42, 1–75. Ukarumpa: Ukarumpa: Summer Institute of Linguistics.

Croft, William. 2022. *Morphosyntax: Constructions of the World's Languages*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316145289.

Curnow, Timothy Jowan. 1999. Towards a cross-linguistic typology of copula constructions. In *Proceedings of the 1999 Conference of the Australian Linguistic society*, vol. 11, 203–300.

Dixon, R. M. W. 2002. *Australian Languages: Their Nature and Development* (Cambridge Language Surveys). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486869.

Dryer, Matthew S. 1992. The Greenbergian Word Order Correlations. *Language*. Linguistic Society of America 68(1). 81–138. https://doi.org/10.2307/416370.

Gong, Liwei & Satoshi Uehara. 2024. Encoding of nominal predication constructions: a typological investigation in verb-initial languages. *Linguistic Typology*. https://doi.org/doi:10.1515/lingty-2023-0035.

Haspelmath, Martin. 2024. Nonverbal clause constructions. To appear in *Language and Linguistics Compass*.

Hengeveld, Kees, Jan Rijkhoff & Anna Siewierska. 2004. Parts-of-speech systems and word order. *Journal of Linguistics*. Cambridge University Press 40(3). 527–570. https://doi.org/10.1017/S0022226704002762.

Katz, Aya. 1996. *Cyclical grammaticalization and the cognitive link between pronoun and copula*. Rice University PhD Thesis.

Piantadosi, Steven T., Harry Tily & Edward Gibson. 2012. The communicative function of ambiguity in language. *Cognition* 122(3). 280–291. https://doi.org/10.1016/j.cognition.2011.10.004.

Shimoji, Michinori. 2008. *A Grammar of Irabu, a Southern Ryukyuan Language*. Canberra: Australian National University PhD Thesis.

Stassen, Leon. 1997. *Intransitive predication*. Oxford: Oxford University Press.

Wasow, Thomas. 2015. Ambiguity Avoidance is Overrated. In Susanne Winkler (ed.), *Language and Communication*, 29–48. Berlin, München, Boston: De Gruyter. https://doi.org/doi:10.1515/9783110403589-003.

Possessors in transitive clauses in Northern Khanty

Nikita Muravyev, University of Hamburg

Grammatical relations in Northern Khanty are said to be directly linked to information-structure (Nikolaeva 2001; Filchenko 2012; É. Kiss 2019). Topical A normally appears in the subject position of a transitive active clause (1). Non-topical A is demoted to oblique via passivization while topical P is promoted to subject (2). Whenever both A and P are topical, a special active subject-object indexing verb form is used, as in (3).

*(1) maša-jen   amp   **λapt-əs***
    Masha-PROP   dog   feed-PST[3SG]
    {Who did Masha feed?} 'Masha fed THE DOG.'

*(2) amp   maša-jen-ən   **λapət-s-a***
    dog   Masha-PROP-LOC   feed-PST-PASS[3SG]
    {Who fed the dog?} 'MASHA fed the dog.'

*(3) maša-jen   λʉwti   **λapət-s-əλλe***
    Masha-PROP   it.ACC   feed-PST-3SG.SO
    {What did Masha do to it?} 'Masha FED it.'

However, a presence of the possessor in the noun phrase of a core argument can have an impact on its coding in the clause leading to coding patterns that may seem somewhat surprising or unexpected. First, as has been observed in (Nikolaeva 2001), if a non-topical P-argument, such as 'eyes' in (4), has a topical possessor (in this case the clause subject), it always triggers object indexing on the verb.

*(4) λuweλa   isa   ośxuλ,   **sɛm-ŋəλ**   tiw   toxi   śi   **tǫt'λ'ə-λ-λe***
    she.DAT   all   interesting eye-DU.POSS3SG here   there FOC   carry-NPST-3SG.SO
    'She is interested in everything, she looks (lit.: moves her eyes) here and there.'

Second, non-topical Ps, such as 'hand' in (5), can hold the subject slot if their possessor stands in an anaphoric relation with a current discourse topic, see e.g. discussion of possessors and switch-reference in (Bárány, Nikolaeva 2019). Here the subject 'wife' from the first clause is not kept in the subject position of the subsequent clause because it is the 3SG possessor (her husband) who is the central participant in the current stretch of discourse.

*(5) **ime-λ**   nox wɛrλə-s   pa   **jɔš-əλ**   jira   **katəλ-s-a***
    wife-POSS.3SG   up wake-PST[3SG]   and   hand-POSS.3SG away   catch-PST-PASS[3SG]
    'Meanwhile his wife woke up and grabbed him by the hand (lit. his hand was grabbed).'

Another feature not mentioned in the literature is possessor promotion in passive constructions from intransitive motion verbs, as in (6) below. Here we observe a striking form-meaning mismatch, as the formal promotion of the goal argument 'house' to subject "hides" the pragmatic promotion of the topical 3SG possessor affected by the entering event, cf. a typological discussion of possessor raising and adversative passives in (Shibatani 1994).

*(6) imuλtijən   **χot-əλ**   χujat-ən   śi   **juχət-s-a***
    suddenly   house-POSS.3SG who.INDEF-LOC FOC   come-PST-PASS[3SG]
    'After he went to sleep, suddenly someone came into his house.'

In this talk, based on the manually annotated examples from an unpublished Northern Khanty corpus (42174 words), I will discuss these and similar coding patterns that result from the presence of a possessor in a clause. I will argue that, since non-pronominal subjects, direct objects, and possessors in Northern Khanty do not exhibit dedicated case morphology, the presence of all three of them requires special disambiguation strategies that rely on principles of communicative efficiency, such as those proposed in (Hawkins 2004).

References

Bárány A., Nikolaeva I. 2019. Possessors in switch-reference. Glossa, 48(1), 1–36.

É. Kiss, K. 2019, Fused grammatical and discourse functions in Ob-Ugric: Case, agreement, passive. — Arbeitspapier 130, 163–174.

Filchenko A. 2012. Continuity of information structuring strategies in Eastern Khanty: Definiteness/topicality. In Pirkko Suihkonen, Bernard Comrie & Valery Solovyev (eds.), Argument Structure and Grammatical Relations: A crosslinguistic typology (Studies in Language Companion Series 126). Amsterdam: John Benjamins Publishing Company, 115–132.

Hawkins J. A. Efficiency and complexity in grammars. – OUP Oxford, 2004.

Nikolaeva I. 2001. Secondary topic as a relation in information structure. In Linguistics, Vol. 39, 1. 1–50.

Shibatani M. 1994. An integrational approach to possessor raising, ethical datives, and adversative passives. *Annual Meeting of the Berkeley Linguistics Society*, vol. 20, no. 1, pp. 461–486.

Abbreviations

ACC — accusative, DAT — dative, DU — dual, FOC — focus, INDEF — indefinite, LOC — locative, NPST — non-past, PASS — passive, POSS — possessive, PROP — proper noun, PST — past, PTCP — participle, SG — singular, SO — subject-object.

The English inferential construction: form and function

This paper deals with a type of focus construction that has received some modest attention over the last three decades, namely the so-called 'inferentials' as in (1). Inferentials consist of a matrix clause with *it* + *be* + a complement clause. While the (non-)referential nature of *it* is a controversial topic (compare e.g. Otake 2002 and Delahunty 1995), most studies assume the subject to be a dummy *it* as it cannot be questioned or substituted by demonstratives.

(1) For a PI, lying is an important job skill. **It's not that I'm inherently dishonest; it's just that sometimes it pays to preserve wriggle room.** (COCA)

Delahunty (1995) argues that inferentials may fulfil various discourse-pragmatic functions depending on the context, but typically express an explanation or reason for a preceding statement. Declerck (1992) essentially considers inferentials to be semantically specificational and explains their usage in terms of an inferred variable and a corresponding value. In example (1), the speaker uses the inferentials to clarify what value (best) satisfies an inferred variable (e.g. a reason, cause, result, etc.).

Based on data from the *Corpus of Contemporary American English* (COCA), the paper investigates the use of inferentials in English. Following Declerck (1992), it delimits inferentials to specificational utterances only (contra Calude & Delahunty 2011; Otake 2002), and proposes different analyses for utterances like (2-5), despite their functional and formal parallels. The juxtaposition between these various constructions is essential in defining inferentials more precisely.

(2) It might be that some remaining supporters of the Archon helped it along. (COCA)
(3) It's just as if he's, like, trolling around for anything. (COCA)
(4) It's not like the suburbs are going to go away. (COCA)
(5) Not that I care. (COCA)

The paper addresses (i) what the form and functions of the English inferential construction are and (ii) how Functional Discourse Grammar (FDG) can help to account for the similarities and differences between the inferential construction and other constructions such as (2-5). In FDG, inferentials so far have been analyzed only at the Interpersonal Level with an emphasis operator and optionally a negation operator on the Communicated Content (Hengeveld & Mackenzie 2008: 106, 2018: 38). This paper proposes an alternative analysis with two (co-indexed) Propositional Contents in a specificational configuration, one of which representing an absent (but implied) variable. With regard to negative inferentials, it is argued that the negation operator goes to the overall specificational configuration at the Representational Level (contra Hengeveld & Mackenzie 2018: 38), as the meaning of the negation can be explained in terms of rejecting a value for the inferred variable.

References

Calude, Andreea S., and Gerald P. Delahunty. (2011). "Inferentials in spoken English". Pragmatics 21.3: 307-340.
Declerck, Renaat. (1992). "The inferential it is that-construction and its congeners". Lingua 87: 203-230.
Delahunty, Gerald P. (1995). "The inferential construction". Pragmatics 5.3: 341-364.
Hengeveld, Kees; Mackenzie, J. Lachlan. (2008). Functional Discourse Grammar: A typologically-based theory of language structure. Oxford: Oxford University Press.
Hengeveld, Kees; Mackenzie, J. Lachlan. (2018). "Negation in Functional Discourse Grammar". In: Keizer, Evelien; Olbertz, Hella (eds.). Recent Developments in Functional Discourse Grammar. Amsterdam/Philadelphia: John Benjamins, 17-45.
Otake, Yoshio. (2002). "Semantics and Functions of the It is that-Construction and the Japanese No da-Construction". MIT Working Papers in Linguistics 43: 143-157.

'(Going) home' in typological perspective

Julia Nintemann (University of Bremen)

"There is no place like home" – This sentiment is also reflected in many languages, as constructions involving the concept of 'home' often adhere to their own rules. The talk focuses on constructions where 'home' is the goal of a movement in comparison to other common nouns and toponyms in the same position. Linguistic research has already acknowledged that toponyms often follow rules distinct from those governing common nouns, especially in constructions with spatial meaning (cf., e.g., Stolz et al. 2014; Stolz & Nintemann 2024). Additionally, Haspelmath (2019: 322) observes that "languages sometimes give special treatment to a diverse set of nouns that denote concepts which are commonly used as spatial landmarks, such as '(one's) house' […]".

In Neverver [Austronesian], for example, common nouns must be marked with the general locative preposition *lon* when acting as the Ground (cf. Talmy 1978) in a Goal construction (1), whereas toponyms are generally zero-marked (1). Furthermore, there is a very restricted set of only three nouns that behave like toponyms in that they are zero-marked, one of them being '(one's) home' ((1).

(1)   Neverver [Austronesian]

a.   Ground = common noun (Barbour 2012: 130)

Lesien	*at-uv*	***lon***	*nokhos.*
Lesien	3REAL:PL-go	**LOC**	garden

'Lesien and them went **to** the garden.'

b.   Ground = toponym (Barbour 2012: 110)

Ale	*nimkhut*	*turien*	*at-uv*	***Ø***	*Vorkha.*
then	man	some	3REAL:PL-go	**Ø**	Vorkha.

'Then, some men went **to** Vorkha.'

c.   Ground = 'home' (Barbour 2012: 341)

Ale,	*i-dum*	*i-vlem*	***Ø***	*aiyem.*
then	3REAL:SG-run	3REAL:SG-come	**Ø**	home

'And then she ran back home.'

Matching Haspelmath's (2019) term 'topo-nouns', the Neverver noun *aiyem* 'home' receives special treatment in spatial constructions by following the rules for toponyms rather than those for common nouns. However, cross-linguistically, this special treatment of 'home' does not automatically imply the same treatment as that given to toponyms. Sometimes, strategies deviating from both common nouns and toponyms are employed when 'home' is the goal of a movement, as, e.g., in German *Ich gehe heim* or Swedish *Jag går hem*, both meaning 'I go home'. In both languages, prepositions are required with toponyms and common nouns in Goal constructions, whereas 'home' is expressed adverbially with an expression that is derived from a noun (German *Heim*N, Swedish *hem*N). The existence of constructions like French *chez moi* 'at/to my place/home' in many languages certainly supports the idea that the concept of 'home' receives special treatment. However, these constructions will not be included at this point.

In this talk, I will explore how 'home' in Goal constructions is expressed cross-linguistically. It will be examined if 'home' in these constructions behaves like common nouns, toponyms, or neither. I will demonstrate that the concept of '(going) home' indeed receives special treatment in languages all over the world. For this purpose, I will examine a sample of 100 languages from different areas and of different affiliations from a functional-typological perspective. The data will primarily be taken from descriptive grammars, complemented by primary sources and input from language experts.

Abbreviations

LOC = general locative preposition, PL = plural, REAL = realis mood, SG = singular

References

Barbour, Julie. 2012. *A grammar of Neverver* (Mouton Grammar Library 60). Berlin, Boston: De Gruyter Mouton.

Haspelmath, Martin. 2019. Differential place marking and differential object marking. *STUF - Language Typology and Universals* 72(3). 313–334.

Stolz, Thomas & Julia Nintemann. 2024. *Special Onymic Grammar in typological perspective: Cross-linguistic data, recurrent patterns, functional explanations* (Studia Typologica 34). Berlin, Boston: De Gruyter Mouton.

Stolz, Thomas, Christel Stolz & Sander Lestrade. 2014. *The Crosslinguistics of Zero-Marking of Spatial Relations* (Studia Typologica 15). Berlin: De Gruyter.

Talmy, Leonard. 1978. Figure and ground in complex sentences. In Joseph Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language: Volume 4: Syntax*, 625–649. Palo Alto/Ca.: Stanford University Press.

Lifting the stepchild out of poverty: Text collections as a complement to grammars and dictionaries

Sebastian Nordhoff, Mandana Seyfeddinipur and Christian Döhler

Berlin-Brandenburg Academy of Sciences and Humanities

Franz Boas established the "Boasian Trilogy" in language documentation and description (Himmelmann 1998), consisting of a grammatical description, a dictionary, and a text collection. All three are necessary to get a comprehensive overview of a language, and they complement each other. While we have good outlets for grammars (eg Comprehensive Grammar Library) and dictionaries (eg Dictionaria), such is not the case for text collections. This means that only few of them are published, and even fewer follow the FAIR principles of findability, accessibility, interoperability, and reusability (Wilkinson 2016).

The project Open Text Collections (`http://opentextcollections.org`) will remedy this by making high quality text collections from endangered languages available in an open interoperable format. Next to providing pdfs or printed books to the communities themselves, this setup will also provide the data in CLDF format (Forkel et al. 2018) for downstream use in NLP applications.

Most reference grammars published today are the result of a language documentation project, often part of authors' dissertation projects. These grammars should be data-driven and accompanied by a corpus in order to facilitate the verification or falsification of the analysis (Nordhoff 2008, Mosel 2012). While countless hours are invested into the structuring and glossing of texts, in many cases, however, these texts are not made available in a reusable way. Linguists tend to have them somewhere on their hard drive, or uploaded to an archive but there is no generally established way of publishing them, at least not in a format which would feed further research downstream (e.g. linguistic typology, corpus-based language description, or NLP). This means that these valuable results of language documentation often fail to be discovered.

Open Text Collections will provide a quality venue for publishing text collections, following the setup established by Language Science Press. The platform is community-driven and aims at being attractive to both data producers (ie language documenters) as well as data users (language communities, typologists, NLP practitioners). For data producers, the platform will provide

rigorous peer review, quality control, and top-notch publishing (pdf and print-on-demand), making sure that the time invested in a text collection will not harm job prospects. For data consumers, different outlets will be available to suit different needs: printed books will be available for communities; a search interface (prototype available at `https://imtvault.org`) will be available for typologists, and all data will be available as CLDF dump for NLP practitioners. By making reuse easy, the research will spread more widely, which in turn is very attractive for the data producers.

As of today, there are 5 regional boards and 40 proposed text collections. This presentation will showcase the platform, its motivations, and its benefits for data producers and consumers.

References

Forkel, Robert, Johann-Mattis List, Simon Greenhill, Christoph Rzymski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping & Russell D. Gray. 2018. Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Sci Data* 5. DOI: 10.1038/sdata.2018.205.

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36. 161–195.

Mosel, Ulrike. 2012. Advances in the accountability of grammatical analysis and description by using regular expressions. *Language Documentation & Conservation Special Publication* 4. 235–250.

Nordhoff, Sebastian. 2008. Electronic reference grammars for typology: Challenges and solutions. *Language Documentation & Conservation* 2(2). 296–324.

Wilkinson, M. et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3. 160018. DOI: 10.1038/sdata.2016.18.

Grammatical gender in the playground: language variation amongst Juǀ'hoan teenagers

Lee J. Pratchett, Christian-Albrechts-Universität zu Kiel

The Kalahari Basin is a hotbed for grammatical gender systems of typologically different kinds. This includes the highly number-sensitive and sex-based systems of Khoekhoe and Naro (Khoe-Kwadi family), and the non-sex-based systems of Juǀ'hoan (Kx'a family) and Taa (Tuu family), which are also only mildly sensitive to number. To this, one must add the iconic "Bantu gender type", e.g., Tswana and Herero, and the pronominal gender systems of European languages like Afrikaans and English.

In the description of Kalahari Basin Area languages, several studies attest to the impressive variation of gender systems in otherwise closely related dialects (see e.g., Kießling 208 for Taa). In previous research, I demonstrate how language contact and language ideologies result in the subtle yet typologically significant innovation of sex-based genders in Juǀ'hoan varieties (Pratchett 2021). Motivated by a desire to understand the spread of such features in younger generations, and to more fully describe language diversity, in this talk we explore new data on the gender systems of Juǀ'hoan teenagers, which reveals striking divergence when compared to older generations. But how stable are these systems, and is it possible to know which languages are having the greatest impact on the evolution of grammatical gender in Juǀ'hoan?

References

Kießling, Roland. 2008. Noun classification in !Xoon. In Ermisch, Sonja (ed.), *Khoisan languages and linguistics: proceedings of the 2nd International Symposium, January 8-12, 2006, Riezlern/ Kleinwalsertal*, 225-248. Cologne: Rüdiger Köppe.

Pratchett, Lee J. 2021. An areal and typological appraisal of gender in Ju, STUF
- *Language Typology and Universals*, 74(2): 279–302

Towards a typological description of future anterior outside Indo-European

Forms of the future anterior, second future, futurum exactum or future perfect (cf. English *I will have done*; the names go back to the French, German, Latin and English traditions, respectively) are widely represented in the languages of the world, although they lack a comprehensive typological description. A reason for this gap is the conception of future anterior either as a trivial absolute-relative temporal form (in the paradigm of Reichenbach 1947 or Comrie 1985) yielding the temporal precedence of the situation with regard to a reference point in the future, or, if applies, as a morphological counterpart of the present perfect shifted into the future temporal frame by the substitution of an auxiliary verb. However as shown in recent works (primarily Penkova 2018, 2019), the forms of the future anterior in different Indo-European languages are characterized by a polyfunctionality that cannot be boiled down to the compositional sum of their structural components. In this respect, the future anterior is not unlike pluperfect, typologically described since Dahl 1985, as a polyfunctional TAM category in its own right. These non-compositional uses, described in Penkova 2018 and 2019 for the languages of Europe (with a special emphasis on Slavic) and in Sitchinava 2022 for Romance, include evidential and modal uses, conditionals, or verificatives ('it is found out that...').

In the paper I investigate the foci of variation of form and non-compositional semantics of future anterior outside Indo-European and in what regard it informs our knowledge on its uses driven from the Indo-European languages. The bulk of languages where future anterior forms were found comes from Africa and Oceania, however these forms are occasionally attested also in other language areas, including the Volga area or the Americas.

From the formal point of view, in many languages the forms expressing "anteriority in the future" are not strictly compositional. If a form has two components, any of them may have another default interpretation outside the context of the future anterior form. Anteriority might be marked by completive (as in Réunion Creole) or perfective (as in Nzime / Bantu A); futureness might be represented by more general irrealis forms (as in Northern Mao / Omotic or Neverver / Oceanic). A class of 'prospective shift markers' (symmetrical to the so-called 'discontinuous past' markers, cf. Plungian, Auwera 2006) can be discerned in some languages, including Dongolawi / Nubian and Hill Mari / Uralic. In the latter, the resulting form develops secondary verificative uses parallel to the ones in Old Slavic. Some Bantu languages feature additional markers of modality that yield future perfects; this pattern has also a Slavic counterpart, namely the South Slavic forms with additional epistemic marker *da.* Languages with dedicated non-compositional markers are found in Sudanic, Chadic, and Austronesian families.

As far as semantic polyfunctionality is concerned, non-Indo-European languages feature counter-factual uses of future anterior (for example in Omotic languages or Tagalog), weakly attested in the European area. Inferentive uses (such as 'He must have seen us'), well described in Europe, find parallels in Alekano / Kainantu–Goroka and Aghul / Northeast Caucasian. Development towards conditional semantics is found in Bantu and Chadic.

The typological data outside the Indo-European languages provides additional material for building a semantic map for the future anterior polyfunctionality, and confirms its significance as a cluster of uses.

References

Comrie. Bernard. 1985. Tense. Cambridge: Cambridge Univ. Press
Dahl, Östen, 1985. Tense and aspect systems. Oxford: Blackwell

Pen'kova, Yana A. 2018. From Retrospective to Prospective: Grammaticalization of Future Anterior in the languages of Europe // Voprosy Jazykoznanija 2: 53–70

Pen'kova, Yana A. 2019. Modal and evidential strategies for future anterior in European languages: Questionnaire and corpus data // Voprosy Jazykoznanija 6: 7–31

Plungian, Vladimir, van der Auwera, Johan 2006 —   Plungian V. A., van der Auwera J. Towards a typology of discontinu-ous past marking. Sprachtypologie und Universalienforschung — Language Typology and Universals 4: 317–349

Reichenbach, Hans, 1947. Elements of symbolic logic. New York: Macmillan, 1947

Sitchinava, Dmitri V. 2022. Polysemy of Future Anteriorin Romance in a typological context // Voprosy Jazykoznanija 4: 48–65.

Habitual Aspect in Dolgan and Evenki
Eugenie Stapert, Kiel University

In this paper I investigate the habitual aspect in Dolgan, a Turkic language spoken in arctic Siberia, on the basis of recently published corpus data.

It is commonly agreed on that the history of the Dolgan language is characterised by extensive contact with Evenki (Tungusic), Nganasan (Samoyedic) and Russian (Slavic) and that these contacts most likely have played an important role in the development of the observed differences between Dolgan and its closest relative Sakha (Artemyev 2001, Pakendorf 2007, Stapert 2013, Däbritz 2022). One of the domains, in which Dolgan differs from Sakha, is its use of the habitual aspect. Earlier studies have suggested that the habitual aspect is used significantly more frequently in Dolgan than in Sakha, that it is employed in a wider range of contexts, and that its specific habitual meaning may be in a process of bleaching (Stapert 2013). Although this hypothesis was formulated on the basis of solid data, the recent publication of two large corpora of Dolgan and Evenki significantly change the scale on which the patterns of use in Dolgan and Evenki can be compared. The aim of this paper is to provide a large-scale comparative study on the quantitative and qualitative aspects of the habitual in Dolgan and Evenki, as well as to evaluate whether the recently published corpus data confirm or modify earlier conclusions.

Sources:

Artemyev, N.M. 2001. Dolganskij jazyk, chast' 2 [The Dolgan language, part 2]. Saint Petersburg: RGPU im. A.I. Gerzena

Däbritz, C.L. 2022. A grammar of Dolgan: A Northern Siberian Turkic language of the Taimyr Peninsula. Leiden/Boston: Brill

Pakendorf, B. 2007. Contact and prehistory of the Sakha (Yakuts): linguistic and genetic perspectives. Leiden: LOT

Stapert, E.L. 2013. Contact-induced change in Dolgan: an investigation into the role of linguistic data for the reconstruction of a people's (pre)history. Leiden: LOT

Däbritz, C.L; Kudryakova, N; Stapert, E.L. 2022. INEL Dolgan Corpus. Version 2.0. Publication date 2022-11-30. https://hdl.handle.net/11022/0000-0007-F9A7-4. Archived at Universität Hamburg. In: The INEL corpora of indigenous Northern Eurasian languages. https://hdl.handle.net/11022/0000-0007-F45A-1.

Däbritz, Chris Lasse & Gusev, Valentin. 2021. INEL Evenki Corpus. Version 1.0. Publication date 2021-12-31. Archived at Universität Hamburg. https://hdl.handle.net/11022/0000-0007-F43C-3. In: The INEL corpora of indigenous Northern Eurasian languages. https://hdl.handle.net/11022/0000-0007-F45A-1

Thomas Stolz & Nataliya Levkovych

Universität Bremen

stolz@uni-bremen.de, levkov@uni-bremen.de

Assigning English loanwords to grammatical genders in Welsh

In the literature on grammatical gender, there is a controversy whether borrowed nouns are assigned a certain gender according to the rules of the replica language (Corbett 1991) or those of the donor language (C. Stolz 2009). Welsh is a two-gender language (masculine vs feminine) which over many centuries has integrated scores of nouns from English – a genderless language. In this talk, we try to determine in what way gender is assigned systematically to English loan nouns. According to a preliminary count on the basis of the *Geiriadur Prifysgol Cymru* (Thomas 1950–2002) there are 49,086 nouns in the Welsh lexicon, of which 70.31% are registered as masculine as opposed to 20.35% which are feminine. The remaining 9.34% are nouns with double gender. If we focus on nouns of English origin only, the results are as follows: of the 1,815 cases 54.49% involve masculine nouns and 18.46% feminine nouns. The share of double gender nouns increases to 27.05%. There is thus a very pronounced quantitative discrepancy between the frequency of genders in the overall lexicon and the English-derived segment thereof. It is especially intriguing to see that borrowings from English can frequently oscillate between genders whereas this possibility is much more limited in the bulk of the Welsh lexicon.

We will test whether the distribution of borrowed nouns over genders is determined by phonological, morphological, or semantic factors. The results will make it possible to situate Welsh in the typology of contact gender that is currently under debate.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Thomas, Richard James (ed.). 1950–2002. *Geiriadur Prifysgol Cymru*. Caerdydd: Gwasg Prifysgol Cymru.

Stolz, Christel. 2009. A different kind of gender problem: Maltese loan-word gender from a typological perspective. In Bernard Comrie et al., *Introducing Maltese Linguistics: Selected papers from the 1st International Conference on Maltese Linguistics, Bremen, 18–20 October, 2007*, 321–353. Amsterdam, Philadelphia: Benjamins.

Comparative constructions in Tungusic: The limits of variation

The paper focuses on comparative constructions in Tungusic languages (Siberia and Northern China). In all Tungusic, except for Evenki, the Parameter (adjective or adverb) in the comparative construction is unmarked or optionally marked (see Alonso de la Fuente 2011). The Standard of comparison can be encoded by different case forms (1); in the paper, I observe all the available encoding strategies, their genealogical and areal distribution, as well as probable scenarios of their spread.

(1) Gorin Nanai (the archive of Aleksandra Putintseva)

Tuliə-du-ni	*xamača=da*	*seasi-ni*	*chopal*	*bi,*
yard-DAT/LOC-3SG	which=EMPH	sound-3SG	all	be.PRS
*burə-**duki**=də*		*ulən.*		
accordion-**ABL**=EMPH		good		

'There is some <u>sound</u> <COMPAREE> in his yard, <u>better</u> <PARAMETER> than **an accordion** <STANDARD>.' (e_tumali: 1935-8-53-87)

Most of Tungusic use the locational strategy in terms of Stassen (2013), which is cross-linguistically the most widespread, i.e. encode the Standard of comparison as Source of motion (from-comparatives), Goal (to-comparatives), or Location (at-comparatives). All three types are attested in Tungusic, from-comparatives are best represented. Some Tungusic use strategies that are more exotic in a cross-linguistic perspective. In Ulcha and Uilta, the Standard of comparison is marked with the instrumental case, Amur Nanai uses a dedicated comparative case form. In the paper, I will discuss how these strategies develop.

The greatest variability in the comparison encoding is attested in the Amur region, in which the density of Tungusic varieties is the highest. More precisely, the Nanaic branch demonstrates the highest number of strategies, see Figure 1. Exactly this branch lacks the ablative marker -*dōki* (except for Gorin Nanai, contacting Ewenic), which is the most common marker of comparison in other Tungusic.

As for the range of forms encoding comparison, they reveal even more variation, see Figure 2. Along with the ablative -*dōki*, several other ablatives are used in this function. At the same time, some markers listed in Figure 2 that synchronically have no Source meaning might be diachronically related to -*dōki* (including the dedicated comparative -*dūj* in Amur Nanai, see Kazama 2024: 376).

Two competing diachronic tendencies can be postulated for Tungusic. The first one is to inherit the pattern (from-comparatives, typical not only for Tungusic, but also for other languages of the area). The second one is to inherit the form, once assigned for the Standard of comparison. I will discuss them in a more general perspective of inheritance vs. borrowing of valency patterns (see Say 2018; Grossman & Witzlack-Makarevich 2019).

References

Alonso de la Fuente, Jose A. 2011. Northern Tungusic*/Vº-tmAr/~*/Cº-d (ï) mAr. *Acta Orientalia Academiae Scientiarum Hungaricae*, 64(2). P.183–200.

Grossman, Eitan, & Alena Witzlack-Makarevich. 2019. Valency and Transitivity in Contact: An Overview. In: Grossman, Eitan, Witzlack-Makarevich, Alena, and Ilja Seržant. (eds.). Valency and transitivity in contact: theoretical and empirical issues. [Special issue]. *Journal of Language Contact*, 12(1). 1–26.

Kazama, Shinjiro. 2024. Nanai. In: Vovin, A., Alonso de la Fuente, J.A., and J. Janhunen (eds.). *The Tungusic Languages*. London / New York: Routledge. P. 364–406.

Say, Sergey. 2018. *Markirovanie aktantov dvuhmestnyh predikatov: predvaritel'nye itogi tipologičeskogo issledovanija* [Argument encoding in bivalent predicates: preliminary results of a cross-linguistic study] In: Say, Sergey (ed.). *Valency classes of bivalent predicates in various languages*. Sankt-Petersburg: ILI RAN. P. 557–616.

Stassen, Leon. 2013. Comparative Constructions. In: Dryer, Matthew S., & Martin Haspelmath (eds.). *WALS Online* (v2020.3) https://doi.org/10.5281/zenodo.7385533 (Available online at http://wals.info/chapter/121, Accessed on 2024-01-11.)

Appendix

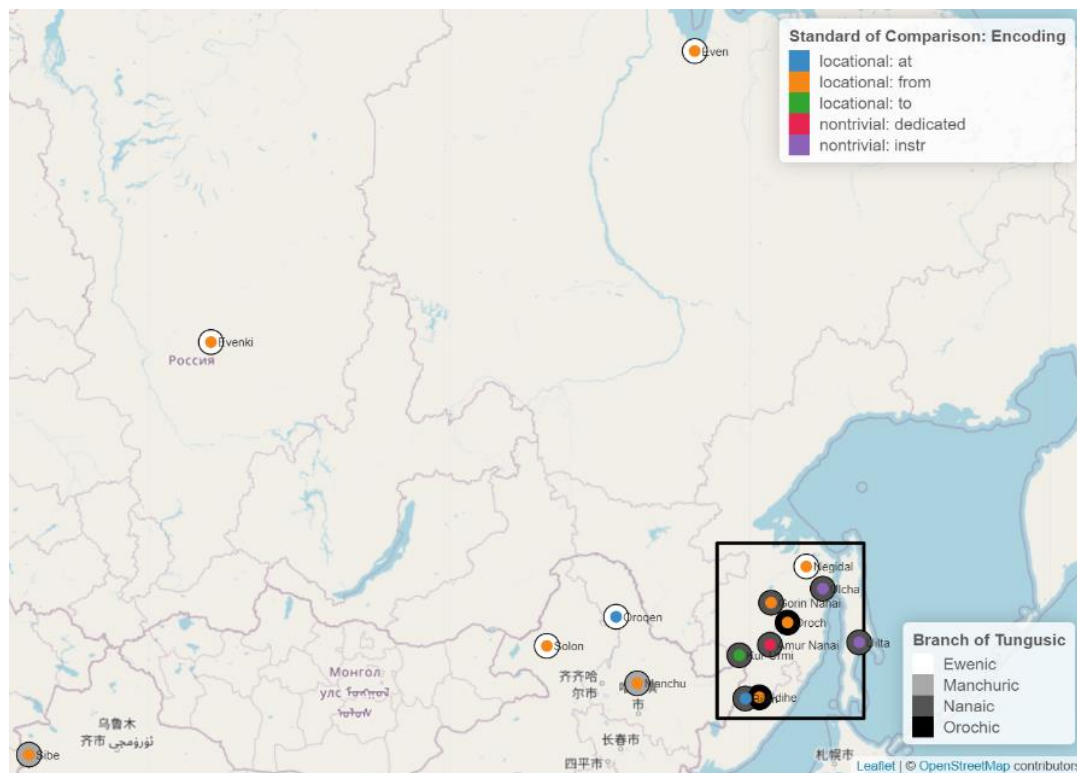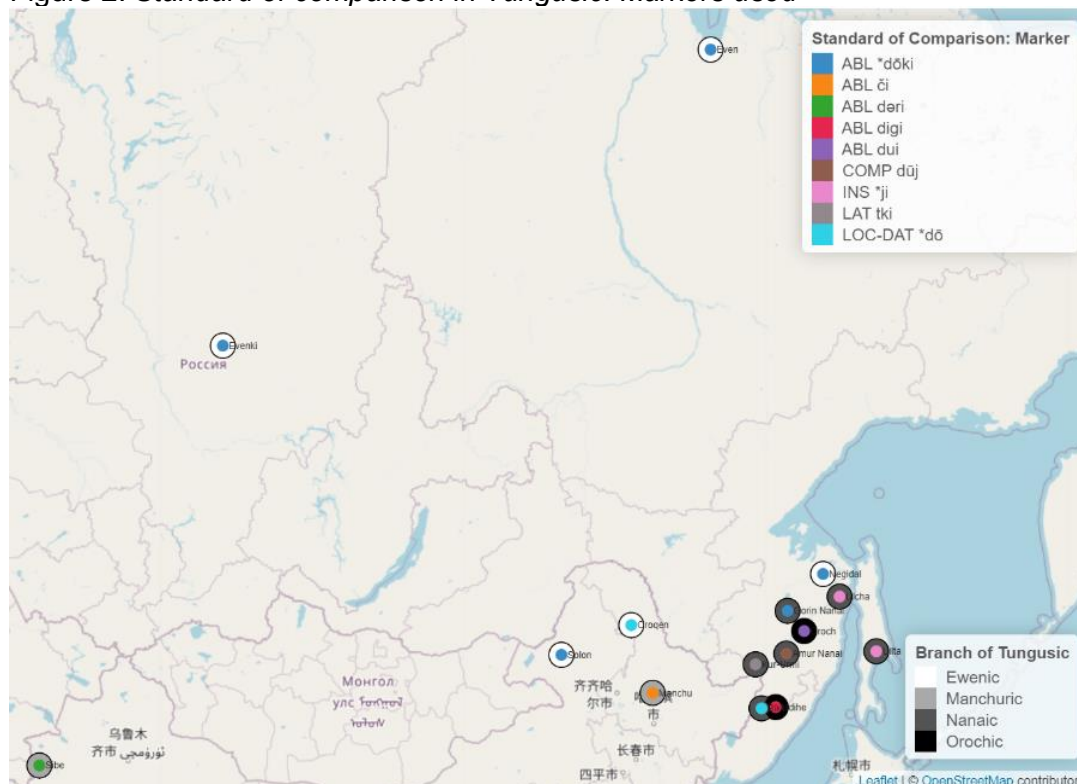Figure 1. *Standard of comparison in Tungusic: Encoding types*

Figure 2. *Standard of comparison in Tungusic: Markers used*

"Speech to which no one says mhmm is an orphan": Continuers and filled pauses in Kambaata conversations

Yvonne Treis (CNRS-LLACAN) in collaboration with Yoseph Yonas Zecharias (University of Gonder) & Teshome Dagne Madebo (Kambaata Zonal Administration)

In our talk, we focus on two salient aspects of interactive language data from the Ethiopian language Kambaata (Cushitic): (i) continuers and (ii) filled pauses. Continuers, also called "backchannels", signal to interlocutors that one is listening and waiting for them to continue (see, e.g., Dingemanse, Liesenfeld & Woensdregt 2022). As the proverb (quoted from Alamu & Alamaayyo 2017: 152) in our title is meant to illustrate: Kambaata speakers expect their interlocutors to engage in back-channeling when they are being told a story.

We adopt a bottom-up approach to Kambaata continuers, which belong to the word class of interjections. We identify them in our recorded texts and distinguish them from other responses (e.g. *ââ* 'yes',[1] *ekku* 'okay, I agree', *ishshi* < Amharic loan 'okay, I agree'). Then we classify them according to their phonetic forms. The following preliminary categories could be established: 1. *mhmm ~ ûhûû*, 2. *ââ ~ âhââ*, 3. *ee*, and, most notably, 4. an ingressive (inhaled) feedback signal [↓] (Eklund 2008). Our paper attempts to determine their distribution and discusses in how far they are exchangeable or expressing subtle differences in the type of feedback they provide.

The second topic of our talk are so-called "filled pauses" (see, e.g., Kosmala & Crible 2022; Kaland & Bardají i Farré 2023). For Kambaata, we distinguish between more *ĩĩ*- and more *ee*-filled pauses at the beginning of turns. After having demonstrated the use of these pause fillers, we attempt to determine whether they are rather signs of fluency (speaker holding the floor, signaling speech continuation to the interlocutor) or disfluency (hesitation, search for words).

The paper is based on recorded, transcribed and translated Kambaata conversations of two speakers each, about such diverse topics as beekeeping, blessing and cursing traditions, and exchanges about picture stories. Altogether the sample texts feature 5 different speakers.

---

[1] The circumflex ^ stands for nasalization.

References

Alamu Banta & Alamaayyo G/Tsiyoon. 2017. *Hambarrichcho yaanata: Kambaatissa-Amaarsa hayyo'ooma yannaakkata [Spices of Mount Hambarrichcho: Kambaata-Amharic proverbs]*. Addis Ababa: Addis Ababa University.

Dingemanse, Mark, Andreas Liesenfeld & Marieke Woensdregt. 2022. Convergent evolution of continuers. In Andrea Ravignani, Rie Asano, Daria Valente, Francesco Ferretti, Stefan Hartmann, Misato Hayashi, Yannick Jadoul, et al. (eds.), *The evolution of language: Proceedings of the Joint Conference on Language Evolution (JCoLE), Kanazawa, Japan, 5-8 September, 2022*, 160–167. Nijmegen: Joint Conference on Language Evolution (JCoLE). https://doi.org/10.17605/osf.io/v6kfn.

Eklund, Robert. 2008. Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association* 38(3). 235–324. https://doi.org/10.1017/S0025100308003563.

Kaland, Constantijn & Maria Bardají i Farré. 2023. Phonetic description of filled pauses as discourse markers in Totoli. In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences, Prague 2023*, 3242–3246. Guarant International.

Kosmala, Loulou & Ludivine Crible. 2022. The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech* 65(1). 216–239. https://doi.org/10.1177/00238309211010862.

Radical typology: on root length and structure cross-linguistically.

Writing at the dawn of modern linguistics in the 19th century, Franz Bopp (1845: 102) distinguished "languages with monosyllabic roots, without the capability of composition... where all is hitherto bare root...," "languages with monosyllabic roots, which are capable of combination," and "languages with dissyllabic verbal roots [...]" What we have here, intertwined with well-known early typological categories corresponding to "isolating", "inflecting", and the "introflecting" Semitic languages, is the outline of a typology of root shapes.

The shape of lexical roots which Bopp incorporated in his typology, in particular their length in terms of syllables, is, in spite of these hopeful beginnings, still one of the empirically and theoretically least explored typological parameters 180 years later.

Here, I will sketch a first systematic stab at exploring this topic. I will provide an overview of root shape as a typological variable, distinguishing three descriptive parameters: (i) *length* (here, in terms of syllables, though potentially also in terms of segments and morae), (ii) *canonicity*, i.e. whether there are strong preferences regarding the shape of these roots or whether they vary considerably, and (iii) *word class specificity*, which captures the fact that in some languages (e.g. Basque) both parameters (i) and (ii) differ notably between nouns and verbs.

The empirical base is a sample of 77 languages, for which all in all thousands of roots have been evaluated. In the (notionally) nominal domain, data are from Urban (2012); for the verbal domain, I evaluate for the first time newly gathered data for 121 verbal concepts (from Tomasello 1992). I will sketch descriptive statistics of this dataset; for instance, regarding parameter (i) (length) the density plot to the right shows that the distribution of mean length of roots is bimodal, with peaks at lengths a little larger than one and two syllables. These are caused by languages with strong canonicity (parameter ii).

Mean length across 77 languages in terms of syllables. Bold: Overall, Solid: Verbs, Dashed: Nouns
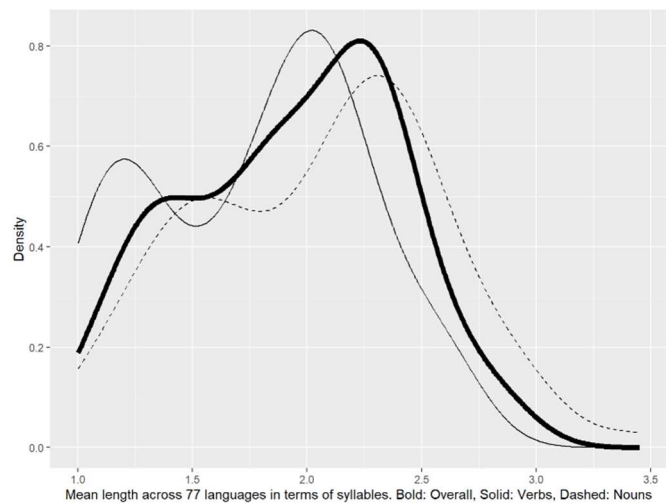
I will also briefly address the typologically relevant questions as to what conditions the behavior of individual languages. I will tie this question to two very different traditions of cross-linguistic work that are pertinent here. First, research in the tradition of linguistic economy and efficiency, which would posit an optimization of root length in light of the phonological inventories (and possibly further factors). With regard to root length, this idea is prominently represented in the work of Prague School typologist Vladími Skalička (e.g. [1946]1979). Second, research in the prosodic typology of languages that posits far-reaching consequences of the rhythmic structure of languages across the phonology, morphology and syntax (Donegan and Stampe 2004). First results of these inquires will be available by the time of the conference.

Bopp, F. 1845. A comparative grammar of the Sanscrit, Zend, Greek, Latin, Lithuanian, Gothic, German, and Sclavonic [sic!] languages. London: Madden and Malcolm.

Donegan, P.J, & Stampe, D. 2004. Rhythm and the synthetic drift of Munda. In Rajendra Singh (ed.), The Yearbook of South Asian Languages and Linguistics 2004, 3-36. Berlin: DeGruyter.

Skalička, Vladímir [1946]1979. Über die Typologie des gesprochenen Chinesisch. In: Peter Hardmann (ed.): Sprachtypologische Studien, 180-197. Braunschweig/Wiesbaden: Friedr. Vieweg & Sohn.

Tomasello, M. 1992. First verbs: a case study of early grammatical development: Cambridge: Cambridge University Press.

Urban, M. 2012. Analyzability and semantic associations in referring expressions: a study in comparative lexicology. Dissertation, Max Planck Institute for Evolutionary Anthropology/Leiden University.

Schriftliche und mündliche Quellen in der Sprachforschung

***

Rainer M. Voigt, FU Berlin

Die traditionelle semitistische Methodik, bei der Beschreibung einer Schriftsprache nach den schriftlichen Quellen vorzugehen, hat sich in der Geschichte der Semitistik bewährt und wird heute bei seit dem 19. Jh. neuentwickelten Schriftsprachen in Äthiopien und Erythräa beibehalten. Muttersprachler werden demgegenüber befragt, wenn es um die Aufnahme von Dialekten geht.

Die jüngste Kritik an dieser Methode von dialektologischer Seite her ist ein Anlass, die beiden deskriptiven Ansätze der Beschreibung einer Literatursprache und der gesprochenen Sprache stärker voneinander abzugrenzen. Der Unterschied betrifft nicht ausschließlich den zwischen Hochsprache und Dialekt, obwohl dies gewöhnlich so gesehen wird. Vielmehr sollten die Linguisten, die bei der Beschreibung einer Schrift- sprache Muttersprachler befragen bzw. ihre eigene sprachliche Kompetenz einbringen, alle Aspekte der gesprochenen Sprache im Auge haben, wie vokalische und konsonan- tische Assimilationen auf lautlicher Ebene, Varianten auf morphologischer Ebene sowie Wortstellung und Satzkonstruktionen auf syntaktischer Ebene, abgesehen von möglichen Regionalismen, die .

Die modernen Beschreibungen z. B. des Amharischen, die ohne Textbelege arbeiten, geben nicht die tatsächlich gesprochene Sprache wieder, sondern die Hoch- sprache, d.h. die Standardsprache der oberen und mittleren sozialen Schichten, wie sie sich in der Hauptstadt entwickelt hat. Die gondarinisch-amharische Hochsprache weicht schon davon ab. Moderne auf Befragung von Muttersprachlern basierende Beschreibungen, die nicht dialektal ausgerichtet sind, zeigen zwei wesentliche Defizite.

– Zum einen wird die reale Vielfalt der gesprochenen Sprache in diatopischer, diastratischer und diaphasischer Hinsicht ignoriert, und
– zum anderen wird nicht erkannt, dass die Standardsprache der Gebildeten nur ein dürftiger Abklatsch der reicheren und elaborierteren Literatursprache ist.

Die Einbeziehung von Muttersprachlern sollte sich weniger auf deren muttersprach- liche Beurteilungen beziehen, sondern zur Aufzeichnung echter Dialoge führen. Dabei geht es primär nicht um dialektale Formen, sondern um die Art und Weise, wie ein Gespräch zwischen Muttersprachlern abläuft, wobei man da in idealer Weise mit Sprachaufnahmen arbeiten sollte. Die so aufgenommene gesprochene Sprache bietet einerseits ein nicht nur stilistisch, sondern auch grammatikalisch und syntaktisch von der Literatursprache abweichendes Bild, das natürlich in eine Sprachbeschreibung einbezogen werden sollte.

Die gesprochene Sprache ist nur ein schwaches Abbild der elaborierten Hoch- und Schriftsprache. Ohne Berücksichtigung der literarischen Tradition einer Sprache ergibt sich ein dürftiges Bild der sprachlichen Wirklichkeit, während die Nichteinbeziehung der gesprochenen, nicht dialektal geprägten Varietäten der Hochsprache ein nur geringfügig defizitäres Bild vermittelt.

On Maltese coordinating constructions and the elements that add complexity

Maike Vorholt
University of Bremen

In Maltese coordinating constructions involving identical prepositions, a preposition can either precede each complement or just the first one (EQUI-p-deletion, cf. Stolz & Ahrens 2017, Borg & Azzopardi-Alexander 1997: 87). In simple binary constructions the usage of only one preposition is the normal case (cf. Vorholt 2022, Stolz & Vorholt in preparation). In more complex constructions, however, as in (1) to (3), a different picture emerges. In these examples, an overt form of the preposition precedes each of the complements.

This talk takes a closer look at what occupies the space between the preposition's slots , i.e. whether there are elements in between the complements, (1), the syntactical weight of the complements, (2), and the number of complements, (3).

(1) element in between [Korpus Malti 3.0, culture3186]

li	*huma*	***mingħajr***	*żmien*	*u*	*forsi*	***mingħajr***	*tweġiba*
SUB	3PL	**without**	time	and	perhaps	**without**	answer
		PREP	COMPL$_1$	CONJ	ADV	**PREP**	COMPL$_2$

'[...] which are timeless and perhaps **without** an answer.'

(2) syntactically heavy complements [Korpus Malti 3.0, academic12]

interessati	***fi-***	*t-trobbija*	*ta-t-tiġieġ*	*u*	***fi-***	*l-produzzjoni*	*tal-bajd*
interested:PL	**in-**	DEF-upbringing	of-DEF-chicken	and	**in-**	DEF-production	of:DEF-eggs
	PREP	COMPL$_1$		CONJ	**PREP**	COMPL$_2$	

'[...] interested **in** raising chickens and egg production.'

(3) more than two complements [Korpus Malti 3.0, news144327]

f' diversi	*partijiet ta' ġisimha*	*fosthom*	***f'***	*żaqqha*	***f'***	*sidirha*	*u*
in various:PL	part:PL of body:3SG.F	among:3PL	**in**	stomach:3SG.F	**in**	chest:3SG.F	and
			PREP	COMPL$_1$	PREP	COMPL$_2$	CONJ

f'	*idha*
in	hand:3SG.F
PREP	COMPL$_3$

'[...] in various parts of her body, including her stomach, her chest and her hand.'

That these factors might influence coordinating constructions was noticed before by Stolz & Ahrens (2017) for a smaller sample of Maltese prepositions (as the "Bremen List of Maltese Prepositions" was not available then, cf. Stolz & Levkovych 2020, Stolz & Vorholt in preparation). However, not only the phenomenon of interest for this talk is still waiting to be analysed quantitatively. As yet, there is no full-blown quantitative account of Maltese prepositions. My talk is meant to pave the way for future work in this domain.

Relying on data from the *Korpus Malti 3.0* (Gatt & Čéplö 2013) it is explored whether the complexity and number of the complements have an impact on the form of the coordinating construction and how extensive this impact might be.

Abbreviations

3	third person	F	feminine	SG	singular
DEF	definite	PL	plural	SUB	suborinator

References

Borg, Albert & Marie Azzopardi-Alexander. 1997. *Maltese.* London: Routledge.

Gatt, Albert & Slavomír Čéplö. 2013. Digital corpora and other electronic resources for Maltese. In Andrew Hardie & Robbie Love (eds.), *Proceedings of the international conference on corpus linguistics*, 96–97. Lancaster: UCREL. www.mlrs.research.um.edu.mt/CQPweb/malti03/.

Stolz, Thomas & Carolin Ahrens. 2017. On prepositional ellipsis and the factors which block its application in Maltese. In Benjamin Saade & Mauro Tosco (eds.), *Advances in Maltese Linguistics*. Berlin: De Gruyter Mouton.

Stolz, Thomas & Nataliya Levkovych. 2020. From variation towards the grammar of Maltese prepositions - first steps. In Slavomír Čéplö & Jaroslav Drobný (eds.), *Maltese linguistics on the Danube*, 199–240. Berlin: De Gruyter.

Stolz, Thomas & Maike Vorholt. in preparation. *The grammar of Maltese prepositions*. Berlin: De Gruyter Mouton.

Vorholt, Maike. 2022. Preposition deletion in Maltese. Talk on occasion of "Triangulation - linguistic topics from the three different perspectives (Wales, Malta, Northwest Germany)". Delmenhorst, Germany, 13th June 2023.

Zwei Arten von Demonstrativa in Muyu

Alexander Zahrer, Universität Münster

Muyu (Trans New Guinea → Ok) wird von etwa 2.000 Sprechern im westlichen Teil von Neuguinea gesprochen (Zahrer 2023). Die Sprache verfügt über ein reichhaltiges System an Demonstrativa, das über 30 lexikalische Elemente umfasst. Ein Demonstrativum wird hier definiert als lexikalisches Element mit einer deiktischen Komponente, unabhängig davon, in welcher syntaktischen Funktion es auftritt (adnominal, pronominal, adverbial, etc.; siehe Diessel 1999). Morphologisch werden die meisten Demonstrativa mit den zwei gebundenen Morphemen *e-* und *o-* gebildet, deren Wahl einen Kontrast zwischen proximalen und distalen Referenten anzeigt.

Dieser Vortrag konzentriert sich auf zwei Paare grundlegender Demonstrativa: *edo/odo* (Set I) und *ege/ogo* (Set II). Obwohl sich beide grob mit 'dieser vs. jener' übersetzen lassen, unterscheiden Sie sich bei näherer Betrachtung in ihrer syntaktischen Distribution:

Funktion	Set I	Set II
determiniert Subjekt / direktes Objekt	ja	ja
Pronomen zu Subjekt / direktem Objekt	ja	ja
Subjekt in Kopulasätzen	ja	nein
determiniert indirektes Objekt	nein	ja
determiniert Adjunkt	nein	ja
als Relativpronomen gebraucht	ja	ja
adverbiale Funktion ('hier', 'dort')	nein	nein

Im Folgenden wird argumentiert, dass der funktionale Unterschied dieser beiden Sets semantisch begründet ist. Set I verweist direkt auf den Referenten, während Demonstrativa aus Set II auf den Raum zeigen, den dieser Referenten einnimmt. Dennoch kann Set II nicht für typische lokal-adverbiale Funktionen verwendet werden.

In einem letzten Schritt untersuchen wir quantitative Unterschiede dieser Sets bzgl. ihres Auftretens in Kernargumenten (Subjekt, Objekt) und gehen der Frage nach, ob Muyu gerade dabei ist, ein Kasus-System (Nominaltiv/Akkusativ) mittels Demonstrativa aufzubauen.

Literatur

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

- Zahrer, Alexander. 2023. *A Grammar of Muyu.* PhD thesis, University of Münster.

Diachrony of optative constructions in Mari

In this talk I investigate the diachrony of five optative constructions in four Mari lects (< Uralic). The data comes from my fieldwork conducted in 2016–2019 in the villages of Chodrayal (Meadow Mari, Volzhsky district, Mari El), Kuznetsovo, and Mikryakovo (Hill Mari, Gornomariysky district). Besides, literary Meadow Mari data were elicited online in 2021 and supplemented by data from textual sources.

The constructions observed in Mari idioms are provided in Table 1:

№	Construction	Verb form used in the construction
1	V-IMP.3 + (RETR)	3SG imperative (jussive)
2	V-INF-(POSS.2SG) + (RETR)	Infinitive
3	V-NPST.1SG + RETR	Non-past 1SG form
4	V-šaš + (RETR)	Debitive form in –šaš
5	V-PRET + be-COND	Conditional form

Table 1. Mari optative constructions

Each of the lects under consideration features multiple constructions from Table 1, e.g. Kuznetsovo Hill Mari has optative constructions 1, 2, 4 and 5, whereas Chodrayal Meadow Mari has constructions 1, 2, 3 and 4.

In addition, the same constructions in different lects show significant variation in their expression of subject. For example, while in literary Meadow Mari construction 4 is only possible with 1SG subjects (1), in Mikryakovo Hill Mari it is used with all subjects.

(1)   LITERARY MEADOW MARI
    *keč   tora   gəč   pel   šinča   dene   **už-šaš**   *əl'e!*
    at_least   distance   from   half   eye   with   see-DEB   RETR
    'I wish I could at least see it from far away with one eye!' (Vasikova 1975).

(2)   LITERARY MEADOW MARI   MIKRYAKOVO HILL MARI
    a. **jür   **lij-šaš**   əl'e!*   b. *jur   **cärnə-šäš**!*
       rain   become-DEB   RETR   rain   stop-DEB
       Intended: 'I wish it rained!'   'I wish the rain stopped!'

In my talk I will show how the dialectal variation observed with these constructions contributes to a diachronic reconstruction of optative constructions.

Constructions 1, 2 and 5 demonstrate typologically expected paths of grammaticalization. For example, the colexification of jussive and optative forms is discussed by Dobrushina (2011), and the development of if-clauses into optatives by Grosz (2012) and others. The use of infinitive in optative constrictions has parallels e.g. in Russian (*Vot by pojekhat' tuda!*).

I will focus on constructions 3 and 4, which demonstrate less common paths of grammaticalization. For example, in construction 3 (henceforth *1sg-optative*) the lexical verb takes an invariable 1SG form that does not agree with the subject in person. I am not aware of parallels to this development in other languages.

(3) LITERARY MEADOW MARI (http://marlamuter.com/muter/en/Search)
    *ija   nun-əm   **kočk-am**   əl'e!*
    devil   they-ACC   eat-NPST.1SG   RETR
    'Damn them!' (lit. 'may devil eat them!')

1SG-optatives are hardly mentioned in the literature on Mari. However, they are present in literary Meadow Mari, Chodrayal Meadow Mari and Mikryakovo Hill Mari, only being absent from Kuznetsovo Hill Mari.

The use of 1SG forms in an optative construction is enigmatic. A possible scenario could be that *1SG-optatives* may stem from some construction introducing direct speech. For example, in Tatar (< Turkic), which had contact influence on Mari, prospective constructions have developed from direct speech constructions with 1SG forms (Nevskaja 2005):

(4) TATAR (own fieldwork)

agač	***awa-m***	***awa-m***	***di-p***	***tor-a***
tree	fall-PRS.1SG	fall-PRS.1SG	say-CVB	stand-PRS.3SG

‘This tree is about to fall’ (lit.: the tree stands saying: “I will fall”).

However, the Mari 1SG-optative contains no additional morphological material such as *dip tora* in Tatar. Deriving optative meaning from a construction like (4) is also problematic. Instead, constructions like (5) may shed light on the development of 1sg-optatives:

(5) LITERARY MARI (http://marlamuter.com/muter/en/Search)

əš	*šižtare*	*vet,*	*keč*	*šinča-m*	***püjal-am***	***əle!***
NEG.AOR.3SG	warn	PTCL	PTCL	eye-ACC	wink-NPST.1SG	RETR

‘She didn’t warn {me}, at least she could have winked her eye!’

Here, two clauses share the same 3SG subject. The predicate in the second clause expresses speaker’s counterfactual preference and the verb is in 1SG. While the combination of a non-past form and a retrospective marker is expected to yield a counterfactual reading, an interesting question is why the verb in this construction is in 1SG. I propose that this is due to a perspective shift (Spronck et al. 2020), i.e. the speaker literally says ‘She didn’t warn me, {if I were her}, I would have winked at least”. The next step is insubordination of the second clause, which develops into optative.

Additional argument in favor of this scenario is that although *1SG-optatives* are ungrammatical in Kuznetsovo Hill Mari, examples such as (5) are licit in this variety. This means that examples like (5) may indeed instantiate the original construction, from which *1SG-optatives* have originated.

Sources

Dobrushina, N. The optative domain in East Caucasian languages. Tense, aspect, modality and finiteness in East Caucasian languages. 2011. Pp. 95--130.

Grosz P. G. On the grammar of optative constructions. PhD Thesis. Massachusetts, 2012.

Nevskaya I. The typology of the prospective in Turkic languages. Sprachtypologie und Universalienforschung-STUF, 58(1), 2005. Pp. 111–123.

Stef Spronck, An Van linden, Caroline Gentens and María Sol Sansiñena. 2020. Perspective persistence and irregular perspective shift: Mismatches in form-function pairings. Functions of Language 27(1): 1–6 [Special issue ‘Notes from the field on perspective-indexing constructions: Irregular shifts and perspective persistence’, edited by Stef Spronck, An Van linden, Caroline Gentens and María Sol Sansiñena (https://benjamins.com/catalog/fol.27.1)]

Vasikova L. P. Modal meanings of future participles in modern Mari //Voprosy marijskogo yazyka. Joshkar-Ola, 1975. Pp. 46–55. (In Russ.)

Spatial Elements in the Tense-Aspect system of the North Halmahera Languages

The North Halmahera languages form a Papuan language family in the North Moluccas (Indonesia). They use an absolute frame of reference, with two axes: seawards vs. landwards and upwards vs. downwards, as well as the two deictic notions 'to the deictic center' (ventive) vs. 'away from the deictic center' (itive) (cf. Yoshida 1980). Several roots that are employed for spatial reference are also used to express aspect and tense.

For Tidore and Pagu, the use of the respective reflexes of itive Proto-North Halmahera (PNH) *hino and ventive PNH *hika for referring to events in the past and the future is reported. Itive *hino opens up a time frame between the time of speaking and a set time in the future (1) or past (2) within which an event happened viz. will happen.

(1) Pagu

o takol tumudiing dewela**-(i)no** ka wo-momik-oka
NRNM hour seven morning-**VENTIVE** only 3sg.m.A-wake.up-LOC.there
'he always wakes up at seven o'clock in the morning'

lit. 'he will wake up between now and seven o'clock but not later'

(Perangin Angin 2018: 96)

(2) Tidore

...uto goka romoi rai **ino** fajato mancari nde bole-bole bato
plant bone one already **VENTIVE** 1sg.f.A work 3nh.here RED-slow just
'after planting the bone my work became easy'

lit. 'between planting the bones and now, my work became easy'

(van Staden 2000: 353)

Ventive *hika opens up a time frame between a set point in the future (3) or past (4) and an later viz. earlier time.

(3) Tidore

Kabata ena=re=ge ona gosimo yuke toma posa-posa **ia** yo-so-dagi ena
'song' 3nh=here=there 3pl old.people first LOC RED-ago **ITIVE** 3n.A.CAUS-NOM.go 3nh
yo-so-buria na-munara
3n.A-CAUS-NOM.guard/leader 3pl.POS-work
'These kabata, the ancestors long ago carried out to lead/accompany their work'

lit. 'These kabata, the ancestors long ago to even longer ago carried out to lead/accompany their work' (van Staden 2000: 353)

(4) Pagu

o hari Jumat**-ika** o ngoi to-ka-isa
NRNM day Friday-**ITIVE** NRNM 1sg 1sg.A-to-landwards
'on Friday or later I will go landwards.'

(Perangin Angin 2018: 97)

In the Mainland North Halmahera group, a suffix *-(o)ka* is used as a perfective or past tense marker (5). For Galela, Shelden (2002) relates this suffix *-ka* to the ventive root *hika, marking an alternative outcome (a change) of an event, leading to perfective reading. In the other Mainland North Halmahera languages, however, the suffix is *-oka* and therefore not homophoneous with reflexes of *hika (*-(i)ka*)

but with the general locative marker *-oka*. The connection between locative and perfective/past tense is less straightforward but it may lie in the use of *-(o)ka* to mark specific points in time (e.g. *last Friday*). When attached to verbs the suffix may have received a completive reading.

(5) Tobelo

i-hi-timono-**oka**
3sg-1sg-old-**PERF**
'I'm already old.'
(Holton 2003: 44)

In my talk, I will give a comparative overview on the usage of spatial elements in North Halmaheran tense-aspect systems, and offer a historical explanation for the phenomenon. I will argue that while the non-spatial use of *-(o)ka* in the Mainland North Halmahera languages goes back to a shared innovation in this group, the non-spatial use of *hika and *hino developed separately in the individual languages.

Literature

Holton, Gary. 2003. *Tobelo* (Languages of the World/Materials 328). Munich: Lincom.

Perangin Angin, Dalan. 2018. *A descriptive grammar of the Pagu language*. The University of Hong Kong dissertation.

Shelden, Deidre. 2002. The Galela particle 'ka': four deictic functions. *Journal of translation and textlinguistics* 14. 53–76.

van Staden, Miriam. 2000. *Tidore: A Linguistic Description of a Language of the North Moluccas*. Universiteit Leiden dissertation.

Yoshida, Shuji. 1980. Folk Orientation in Halmahera with Special Reference to Insular Southeast Asia. *Senri Ethnological Studies* 7. 19–88.