

## Table of Contents

### **Jana Bajorat**

Optional ergative marking in Ika (Chibchan, Colombia) - Corpus evidence ..... 3

### **Silvia Ballarè, Caterina Mauri, and Andrea Sansò**

Object encoding in spoken language data and antipassives ..... 6

### **Claudine Chamoreau and Natalia Cáceres Arandia**

Optional case marking of post-verbal arguments in Pesh (Chibchan) ..... 8

### **Eberhard Gade, Johannes Helmbrecht, Christian Klöckner, Moritz Weiser, and Benedikt Weiss**

Can frequency explain the asymmetrical coding of proper names compared to common nouns regarding different syntactic roles? ..... 9

### **Dina El Zarka and Sandra Ziagos**

Gender agreement in South-Iranian Arabic – precursors for the loss of grammatical gender due to language contact ..... 12

### **Sofia Ganieva and Karina Sheifer**

Itelmen inverse system in a typological perspective ..... 14

### **Charlotte Hemmings, Erik Zobel and Mary Dalrymple**

A corpus-based analysis of grammatical relations in Enggano ..... 17

### **Linda Konnerth**

(Avoiding) Overt reference to speech act participant objects: Evidence from Trans-Himalayan diachrony and spoken language corpora ..... 20

### **Åshild Næss**

Voice choice in Äiwoo discourse ..... 22

### **Maria Ovsjannikova and Sergey Say**

Valency-encoding devices in a spoken Northeastern Neo-Aramaic corpus ..... 24

### **Pavel Ozerov**

DCM in V-final languages as local communicative instructions: ergative(?) in Anal Naga ..... 26

### **Linn Iren Sjøanes Rødvand**

Post-verbal subjects in an SV language: non-agentivity and discourse structure ..... 29

### **Henrik Rosenkvist**

Non-standard features in spoken Germanic language varieties – the cases of null subjects and negative concord ..... 31

### **Stefan Schnell, Frank Seifart, Nils N. Schibor, and Geoff Haig**

Do difficulties of reference production differ across grammatical relations? An examination of the distribution of hesitation pauses in eight diverse languages ..... 32

### **Jozina Vander Klok**

Voice type and argument realization of applicatives in Javanese discourse ..... 37

**Eline Visser**

Factors in differential subject marking in Uruangnirin: functional, structural and social..... 40

**Maria Vollmer**

A corpus-based study of flexible word order, language contact and language change in Warlpiri  
..... 43

## Optional ergative marking in Ika (Chibchan, Colombia) - Corpus evidence

Jana Bajorat

Humboldt-Universität zu Berlin

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) characterizes 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. SØV, OSV, SVO, 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 (SØV, 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) a. Pedru Juan pas-ʉn nug-in.  
Pedro Juan hit-IPFV AUX.NONEGO-DECL  
'Pedro is hitting Juan.'
- b. Juan Pedru pas-ʉn nug-in.  
Juan Pedro hit-IPFV AUX.NONEGO-DECL  
'Juan is hitting Pedro.'
- c. Juan Pedru=se" pas-ʉn nug-in.  
Juan Pedru=ERG hit-IPFV AUX.NONEGO-DECL  
'Pedro is hitting Juan.'

- d. Pedru pas-ʉn nug-in.  
Pedru hit-IPFV AUX.NONEGO-DECL  
'He/she/it is hitting Pedro.'
- e. Pedru=se" pas-ʉn nug-in.  
Pedru=ERG hit-IPFV AUX.NONEGO-DECL  
'Pedro is hitting him/her/it.'
- (2) a. inʉ(=se") mʉkʉ a"chuk-w-a-o?  
who=ERG cloth wash-EGO-?-Q  
'Who washed the clothes?'
- b. Maria(=se") mʉkʉ a"chuk-u-w-in.  
Maria(=ERG) cloth wash-AUX-EGO-DECL  
'Maria washed the clothes.'

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.

Initial results confirm the above made assumptions for the ergative marking distribution pattern, namely that ergative marking occurs obligatorily in non-canonical argument orders (SØV is the most frequent pattern), and that, in canonical argument orders, the occurrence of the ergative marker signals prominence or focus.

## References

- DANE. (2021). *Resultados del censo nacional de población y vivienda 2018: Pueblos arhuaco, kankuamo, y wiwa*. Retrieved 19.01.2023, from <https://www.dane.gov.co/files/investigaciones/boletines/grupos-etnicos/presentacion-pueblos-arhuaco-kankuamo-wiwa-CNPV2018.pdf>
- Frank, P. S. (1985). *A grammar of ika* (Dissertation). University of Pennsylvania.
- 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, María 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

## Object encoding in spoken language data and antipassives

Silvia Ballarè\*, Caterina Mauri\*, Andrea Sansò°

\* Università di Bologna, ° Università dell'Insubria

Research on antipassives has long recognized some features of these constructions for which various functional and formal explanations have been proposed (see Sansò 2018 for a survey). In almost half of the languages with an antipassive, for instance, the antipassive is lexically restricted, i.e. it is only possible with a subset of transitive verbs (see Polinsky 2013; Heaton 2017: 161). Cases of lexically restricted antipassives include Eton (van de Velde 2008: 129), Hup (Epps 2005: 405-407), and Ainu (Bugaeva 2021), among others. Typically, in these languages the antipassive is possible with verbs of ingestion, interaction, communication, or traditional activities, and is incompatible with action verbs. A related feature is the correlation between antipassives and imperfective states of affairs: if an antipassive construction can have a perfective interpretation, it must also have an imperfective one (Polinsky 2017: 215-216). While the latter universal tendency is generally explained in functional-typological terms as an effect of imperfectivity on the identifiability of object arguments, the frequent lexical restrictions on antipassive derivation have fostered lexicalist approaches to this type of constructions (see Polinsky 2017).

In diachronic-typological terms, these features have been explained as a case of persistence of features of the source constructions from which antipassives emerge through grammaticalization. Some of these sources, for instance, imply an element that automatically triggers a non-specific (and thus, less perfective) configuration: this is what happens when generic objects such as 'thing(s)' are grammaticalized into antipassive markers (Sansò 2017, 2018).

The aim of this paper is to test the hypothesis whether these features of antipassives correlate with tendencies with respect to object encoding and object deletion in spoken language. We will do so by performing a quantitative analysis of object encoding strategies in a corpus of spoken Italian (The KIParla corpus, cf. Mauri et al. 2019, Ballarè et al. 2022, ca. 1.200.000 words). Italian is a language that has no dedicated antipassive construction, and in which the only strategy to demote or defocus an object argument is by deleting it altogether or by coding it as generic (e.g. through general nouns such as *cose* 'things', *roba/robe* 'stuff', or *gente* 'people').

We have extracted the syntactic patterns of object encoding with the transitive verbs that occur more than 50 times in the corpus, and we have classified these verbs into different semantic classes broadly following Levin's (1993) classification. Moreover, we have classified these verbs in terms of object predictability, defined, in the line of Haspelmath (2021: 624), as the degree to which a given argument type can be expected to fill the object position of a given verb. Finally, we have tagged the occurrences of the verbs in terms of pragmatic accessibility of the object in the context (in the sense of Ariel 1988). We thus consider as a dependent variable the object realization (with three values:  $\emptyset$ , generic object, specific object), and as predictors (i) the verb lexeme, (ii) the verb semantic class, (iii) the semantic predictability of the object, (iv) and the pragmatic accessibility of the object. The analysis shows that both object deletion and generic objects in spoken Italian:

- (i) are significantly more frequent with the same verb classes that allow the antipassive derivation more straightforwardly in languages with lexically restricted antipassives;
- (ii) tend to correlate with argument type predictability.

Moreover, object deletion is significantly higher when the referent is highly accessible in the context. The results of the survey thus suggest that typological tendencies are deeply rooted in language usage and that most of the diversity characterizing antipassives across languages can find a sound explanation by combining diachronic insights with the analysis of spoken data.

## References

- Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24: 65-87.
- Ballarè, Silvia, Eugenio Gorla & Caterina Mauri. 2022. *Italiano parlato e variazione linguistica. Teoria e prassi nella costruzione del corpus KIParla*. Bologna: Pàtron.
- Bugaeva, Anna. 2021. Unspecified participant. A case of antipassive in Ainu. In Katarzyna Janic & Alena Witzlack-Makarevich (eds.), *Antipassive. Typology, diachrony, and related constructions*, 213-246. Amsterdam: John Benjamins.
- Epps, Patience. 2005. *A grammar of Hup*. Berlin: Mouton de Gruyter.
- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57: 605-633.
- Heaton, Raina. 2017. *A typology of antipassives, with special reference to Mayan*. PhD Dissertation, University of Hawai'i at Mānoa.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.
- Mauri, Caterina, Silvia Ballarè, Eugenio Gorla, Massimo Cerruti & Francesco Suriano. 2019. KIParla corpus: a new resource for spoken Italian. In Raffaella Bernardi, Roberto Navigli, Giovanni Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it*. (URL: <http://ceur-ws.org/Vol-2481/paper45.pdf>)
- Polinsky, Maria. 2013. Antipassive constructions. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures online*. Leipzig: The Max Planck Institute for Evolutionary Anthropology. (URL: <https://wals.info/chapter/108>)
- Polinsky, Maria. 2017. Antipassive. In Jessica Coon, Diane Massam & Lisa deMena Travis (eds.), *The Oxford handbook of ergativity*, 307-331. Oxford: Oxford University Press.
- Sansò, Andrea. 2017. Where do antipassive constructions come from? A study in diachronic typology. *Diachronica* 34: 175-218.
- Sansò, Andrea. 2018. Explaining the diversity of antipassives: Formal grammar vs. (diachronic) typology. *Language and Linguistics Compass* 12(6): 1-22. (URL: <https://doi.org/10.1111/lnc3.12277>)
- Van de Velde, Mark L. O. 2008. *A grammar of Eton*. Berlin: Mouton de Gruyter.





## Can frequency explain the asymmetrical coding of proper names compared to common nouns regarding different syntactic roles?

Eberhard Gade, Johannes Helmbrecht, Christian Klöckner, Moritz Weiser, Benedikt Weiss  
University of Regensburg

Recent research on the grammatical properties of proper names has shown that proper names are often coded differently compared to common nouns (cf. e.g., Handschuh 2017, Dammel & Handschuh (eds.) 2019, Helmbrecht et al. (eds.) 2017, Helmbrecht 2020, Stolz et al. 2014, Stolz et al. 2017, and others). In particular in dependent marking languages, it has been observed that grammatical relations may be coded differently if proper names are involved. This concerns e.g., specific case markers, case allomorphy, as well as different case categories and even different alignment types. For instance, in Dyirbal (Pama-Nyungan, Australia) proper names in O function receive an accusative case marker, while common nouns have zero (absolutive) which results in a tripartite alignment versus an accusative pattern with pronouns and ergative pattern with common nouns. In addition, the dative case and the comitative case are morphologically more complex with person names than with common nouns (cf. Dixon 2022:69-82). Differential coding like the one in Dyirbal is not rare and demand some explanation.

In the proposed paper, we would like to answer the question, if the text frequency of proper names (across different syntactic roles and compared against NPs with common nouns) could be the basis for an explanation of the asymmetrical coding of proper names of the type illustrated with Dyirbal. The hypothesis we take as starting point is the idea that the low frequency of proper names (compared to common nouns) in discourse is the reason that proper names are generally more marked than common nouns. Behind this hypothesis is the concept of typological markedness (cf. Croft 1990, 2001) and the role of token-frequency in explaining asymmetrical marking (cf. e.g., Haspelmath 2019, 2021).

We compiled a corpus of spoken German (with variation according to region and conversation type) and annotated all proper names according to name type (person names and its subcategories, place names and its subcategories, and others), structural complexity (with or without article, or lexical modifier), and syntactic role (subject, direct object, indirect object, prepositional object, adjuncts). In addition, we will supplement the annotation of the corpora in Multi-Cast (cf. Haig & Schnell 2022) in order to check if the frequency distribution in the German corpus matches the frequencies in spoken corpora of other languages.

For the workshop, we will restrict our investigation to person names. The following corpus studies will be presented:

- 1) Global text frequencies of pronouns, common nouns and proper names (all types) in spoken German and spoken corpora of other languages.
- 2) Frequency distribution of person names (with and without determiners) across the different syntactic roles vis-à-vis common nouns and pronouns.
- 3) Frequency of person names vis-à-vis lexical NPs that refer to individual human referents.

Our study confirms the expectation that proper names are much less frequent than common nouns and personal pronouns (PRO > CN > PN) and can thus be considered marked compared to

the other referential expressions. This holds also for all syntactic roles, but to different degrees. The marked status of proper names could explain the asymmetrical coding of proper names in languages like Dyrbal, Sinyar (Central Sudanic; cf. Boyeldieu 2019), and others that will be presented likewise. The match between low frequency of proper names in text corpora and the structural markedness of proper names in a number of languages is, however, by far not perfect. This is, e.g., the case in spoken German. We will discuss these problematic cases in the light of various theoretical claims that have been postulated in the literature. Firstly, it has been claimed that person names are the prototypical names and, in addition, the prototypical nouns because of their lack of marking, i.e., their unmarkedness (cf. Van Langendonck 2007:171-2). Secondly, it has been claimed that proper names tend to avoid or reduce inflectional morphology, because this increases the recognizability of names. This processing principle has been brought up by Nübling (2012, 2017) under the notion of “onymic schema constancy”. We will discuss the impact of our findings on these postulates.

## References

- Boyeldieu, Pascal. 2019. Proper names and case markers in Sinyar (Chad/Sudan). *STUF - Language Typology and Universals* 72. 467-504.
- Croft, William. 1990. *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, William. 2001. *Radical Construction Grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Dammel, Antje & Handschuh, Corinna (eds.). 2019. *Grammar of Names*. (Special Issue: *Grammar of Names*) *Sprachtypologie und Universalienforschung* STUF 72(4).
- Dixon, R. M. W. 2022. *A New Grammar of Dyrbal*. Oxford: Oxford University Press. 496pp.
- Haig, Geoffrey & Schnell, Stefan (eds.). 2022. Multi-CAST: Multilingual corpus of annotated spoken texts. Version 2211. Bamberg: University of Bamberg. ([multicast.aspra.uni-bamberg.de/](http://multicast.aspra.uni-bamberg.de/))
- Handschuh, Corinna 2017. Nominal category marking on personal names. A typological study of case and definiteness. In: *Folia Linguistica* 51(2):483-504.
- Haspelmath, Martin 2019. Differential place marking and differential object marking. In: *Sprachtypologie und Universalienforschung* STUF 72,3:313-334.
- Haspelmath, Martin 2021. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. In *J. Linguistics* 57:605–633.
- Helmbrecht, Johannes & Nübling, Damaris & Schlücker, Barbara (eds.). 2017. *Namengrammatik*. (Special Issue) *Linguistische Berichte* 23.
- Helmbrecht, Johannes 2020. Form and function of personal names: dimensions of the morphosyntactic diversity. In: Levkovych, Natalia & Nintemann, Julia (eds.). *Aspects of the grammar of names. Empirical case studies and theoretical topics*. (LINCOM Studies in Language Typology 33). 1-25. München: LINCOM.
- Nübling, Damaris. 2012. Auf dem Wege zu Nicht-Flektierbaren: Die Deflexion der deutschen Eigennamen diachron und synchron. In: Rothstein, Björn (ed.). *Nicht-flektierende Wortarten*. 224–246. Berlin, New York: Mouton de Gruyter.
- Nübling, Damaris. 2017. The growing distance between proper names and common nouns in German. On the development of onymic schema constancy. In: Tanja Ackermann & Barbara

- Schlücker (eds.). *The Morphosyntax of Proper Names*. (Special issue). *Folia Linguistica* 51(2):341-367.
- Stolz, Thomas & Levkovych, Natalia & Urdze, Aina 2017. Die Grammatik der Toponyme als typologisches Forschungsfeld. In: Helmbrecht et al. (eds.). *Namengrammatik*. (Special Issue) *Linguistische Berichte* 23:121-146.
- Stolz, Thomas & Sander, Lestrade & Stolz, Christel. 2014. *The cross-linguistics of zero-marking of spatial relations*. (*Studia Typologica* 15). Berlin: Mouton de Gruyter.
- Van Langendonck, Willy. 2007. *Theory and typology of proper names*. Berlin, New York: Mouton de Gruyter.

# Gender agreement in South-Iranian Arabic – precursors for the loss of grammatical gender due to language contact

Dina El Zarka & Sandra Ziagos  
University of 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 (*bint* 'girl', *ʔumm* 'mother'), paired body parts (*ʔajn* 'eye', *jad* 'hand') or some unique referents (*ʔams* 'sun', *ʔard* 'earth') 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 suffix *-i* denoting origin 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. These varieties 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 is to examine under which circumstances gender agreement is not applied by speakers in discourse.

Our data were collected from two South-Iranian provinces in a documentation project funded by the Austrian Science foundations.<sup>1</sup> The data base for the study consists of 10 socio-linguistic interviews or conversations from 6 villages (approx. 5 hours of recording). We annotate factors known to affect agreement such as the order of controller and target and the distance between them, the type of target (attributive/predicate nominal, verbal predicate, pronoun) and the agreement domain as well as semantico-pragmatic conditions such as animacy, definiteness/specificity, agency and topicality. Sociolinguistic variables are speaker age, speaker gender and the density of Arab population. We apply logistic regression analysis to predict the agreement behavior of the targets from the structural and sociolinguistic variables to identify the significant factors for gender agreement. This furthermore allows us to check our data against the predictions of the agreement hierarchy (Corbett 1991, 2006; Comrie 1975) to see how cross-linguistically established conditions affect the ongoing loss of a gender distinction due to language contact.

---

<sup>1</sup> <https://south-iranian-arabic.uni-graz.at/>

A preliminary analysis of 220 targets shows that approx. 80% show feminine singular agreement. A first educated guess suggests that at least humanness and target-controller order are important semantic/structural factors. Furthermore, agreement in the verb *tʃa:n* ‘to be’, which can fulfill different functions (tense marker, existential) that are largely equivalent to those of the verb *bu:dan* ‘to be’ in Persian, seems to be more neutralized more frequently.

The following example illustrates variable agreement within and across speakers.

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

*jaʕni ħorma tɪ-ddaxxalʔ tʃa:n mʊsʕi:be tʃa:n*  
 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)

*ɪl=ʊmm ... ħa:de ʃʊʊʊl-ha maʕlu:m tʃa:n=u*  
 DEF=mother DEM.SG.M work-3SG.F known be.3SG.M.PFV=and  
*t-giff tʃa:n-at ɪl=ɪħwɑ:j*  
 3SG.F-sweep.IPFV be.PFV-3SG.F DEF=house.PL  
 ‘The mother... this was her work, it was clear, she used to sweep the house.’

(3) (afb-bja-ci-si\_001\_100)

*w=ʊmm-l ma: xalla:-ni ɡɑ:m 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.
- 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.

## Itelmen inverse system in a typological perspective

Sofia Ganieva & Karina Sheifer

Itelmen is considered the only Chukotko-Kamchatkan language lacking an ergative construction. The Itelmen case-marking strategy has been characterized as neutral (Bobaljik, Wurmbrand 2002: 4), (Volodin 2012: 96). The subject of an intransitive verb (S), the subject of a transitive verb (A), and the object of a transitive verb (P) are absolutive, i.e. unmarked. In some rare cases, though, the agent of a transitive verb can be marked with locative. If so, the transitive finite verb, on its part, acquires the prefix *n-*, which was previously described as passive (Wolodin, Georg 1999: 163). In this paper, locative agent marking is revised as non-paradigmatic ergativity and passive marking as inverse system. The study is based on the Itelmen corpus<sup>2</sup> developed by the authors of this paper. The assignment of semantic roles in Itelmen is defined by word order and nominal hierarchy proposed in this research. If the subject and object of a transitive verb coincide in person and number as in (1), the interpretation of semantic roles will solely depend on basic SO word order (Volodin 1974: 15).

- (1) a. *isx-∅*      *∅-aŋčp-nen*      *peč-∅*  
father-ABS SBJ3SG-teach-OBJ3SG son-ABS  
'The father taught the son'.  
b. *peč-∅*      *∅-aŋčp-nen*      *isx-∅*  
son-ABS SBJ3SG-teach-OBJ3SG father-ABS  
'The son taught the father'.

If two arguments of a verb are represented with the items of different hierarchical levels (2), the higher one will possess a more significant semantic role, regardless SO word order.

- (2) personal and possessive pronouns, determiner *mniŋ* 'all' > nouns (incl. proper names) and demonstratives > indefinite pronoun *ke* 'someone'

Thus, in (3), the nominal object *isx-∅* 'father-ABS' precedes the pronominal subject *ina-∅* 'he-ABS'.

- (3) *isx-∅*      *ina-∅*      *χiqe k-ŋinu-kazu-in*  
father-ABS he-ABS bad CVB.NARR-feed-IPFV-CVB.NARR  
'He fed his father poorly'. (#'His father fed him poorly'.)

If both arguments have the same hierarchical level, the distribution of semantic roles will be defined by SO word order as in (1).

When either the hierarchy (4) or SO word order (5) is violated, the subject is marked with locative-ergative and the verb (if it is finite) acquires prefix *n-*, which we believe to be an inverse marker. The narrational converb, in its turn, used as an independent predicate lacks it (5).

---

<sup>2</sup> <https://itelmen.corpora.minlang.site/>

- (4) *pʰx-eʔnk*                      *kiza-∅*              *n-ʰama-xen*  
 people-LOC/ERG.PL    you-ABS              INV-kill-OBJ2SG.INV  
 ‘People will kill you’. (*pʰxeʔnk* ‘people’ < *kiza* ‘you’)
- (5) *k-ʰinnu-ʔin*                                      *Anaqa-čχ*              *laχsχ-enk*  
 CVB.NARR-feed-CVB.NARR    Angake-DIM              mother-LOC/ERG  
 ‘Mother fed Angake’. (*Anaqačχ* ‘Angake’ = O, *laχsχenk* ‘mother’ = S)

When only one argument of transitive predicate is overt, it is marked with locative-ergative iff it is an agent (6). In absence of marking it should be interpreted as patient (7).

- (6) *Siʔrim-enk*                      *k-əʰčku-ʔin*  
 Sirim-LOC/ERG              CVB.NARR-see-CVB.NARR  
 ‘Sirim saw [him]’. (#‘[Someone] saw Sirim’.)
- (7) *silatumx-∅*                      *∅-enxsu-qazu-z-nen*  
 older.sister-ABS              SBJ3SG-search-IPFV-PRES-OBJ3SG  
 ‘[She] is looking for an older sister’. (#‘The older sister is looking for [someone]’.)

Single arguments of intransitive predicates are never marked with locative-ergative (which is typical of ergative case marking), as can be seen from comparison of (9) and (10).

- (8) *ke-∅*                                      *maʔ*                                      *kkuʰuʰa-te-s-č*  
 someone-ABS              somewhere              lament-ITER-PRES-SBJ3SG  
 ‘Someone somewhere is lamenting (vi)’.
- (9) *katx* *k-enk*                                      *k-txunk-qzu-knen*                                      *xanke*  
 as.if    someone-LOC/ERG              CVB.NARR-drag-IPFV-CVB.NARR              up  
 ‘As if someone was dragging [him] up (vt)’.

All the above presented data suggest that the Itelmen locative case marking can be referred to as non-paradigmatic ergativity (commonly found, for example, in Tibeto-Burman languages, see LaPolla 1995). The Itelmen ergative marking is non-systemic and is only used for disambiguation of two potential agents, that is, to indicate a non-prototypical agent (i. e. an agent that is lower than a patient in nominal hierarchy or an agent that is located in an unexpected place of a sentence).

In the Chukchi-Koryak languages, of the same family as Itelmen, ergative marking is systemic and is applied regardless of inverse context. There the ergative case is formally identical to instrumental. In some isolated cases, the Itelmen ergative is also expressed with instrumental, which might be the result of contact interference with Koryak. Chukchi has three morphological markers of inverse alignments, *ne-*, *ine-* and *-tku* (Dunn 1999: 183). The Chukchi-Koryak inverse markers *ine-* and *ne* are to be compared with the single Itelmen marker *n-*. The Chukchi inverse system is primarily based on the person hierarchy (Dunn 1999: 182), whereas the Itelmen one is roughly based on the definiteness scale.

## References

- Bobaljik J. D. & Wurmbrand S. 2002. Notes on Agreement in Itelmen, volume 1.
- Dunn M. 1999. A Grammar of Chukchi. Ph.D. thesis, Australian National University. 386 P.
- LaPolla R. J. 1995. Ergative marking in Tibeto-Burman // Nishi, Matisoff and Nagano (eds.) (1995). PP. 189–228.
- Volodin A. P. 1974. On the question of the ergative construction of a sentence (on the material of the Itelmen language) // Questions of Linguistics No. 1. PP. 14–22.
- Volodin A. P. 2012. On the functional-semantic field of voice // From meaning to form, from form to meaning. Moscow. Language of Slavic cultures.
- Wolodin A. P. & Georg St. 1999. Die Itelmenische Sprache. Grammatik und Texte. Wiesbaden. 323 S.



## A corpus-based analysis of grammatical relations in Enggano

Charlotte Hemmings<sup>i</sup>, Erik Zobel<sup>j</sup> and Mary Dalrymple<sup>i</sup>  
University of Oxford<sup>i</sup> and Independent Researcher<sup>j</sup>

This paper presents a diachronic study of the usage of different constructions to express grammatical relations in Enggano (Austronesian, Sumatra, Indonesia) by comparing an older corpus collected in 1930s (Kähler 1955, 1957, 1958, 1960a, b, 1961, 1962, 1964, 1975) with a contemporary corpus collected as part of an ongoing documentation project since 2018.

Unlike other Western Austronesian languages, Enggano does not possess a symmetrical voice system (Riesberg 2014). Instead, erstwhile voice morphology surfaces in nominalisations and in the verbal marker *bu-* (from PAN *\*-um-* (Edwards 2015)) which alternates with bare verbal stems (potentially derived from a dependent undergoer voice form) and verbs marked with *ki-* (an innovative form cognate with the Nias relativiser *si=* (Brown 2001)).

- (1) a. ka-**bu**-pudu-ha      epaE e'ana      *bu*-form  
3.NOM-**bu**-kill-EMPH    child DEM.MED  
'and he killed the child' (Kähler 1955:90)
- b. kea-ba'a      i-pudu      e-koyo e'ana      bare form  
NEG-INTENSIVE 3.ERG-kill    DIR-pig DEM.MED  
'He didn't kill the pig' (Kähler 1940:101)
- c. 'o'o **ki**-pudu e-koyo      e'ana      *ki*-form  
2SG FOC-kill    DIR-pig      DEM  
'You killed the pig' (Kähler 1940: 113)

Like a number of languages in Sumatra and Sulawesi, Enggano has a system of person-marking. A first set of markers co-occur with *bu-* verbs (1a), and a second set are used with bare verbs (1b). These are labelled NOM and ERG on the basis of cognate sets in other languages, but both co-reference S/A in an accusatively-aligned system. *Ki-* verbs never take agreement. Instead subjects are expressed using free pronouns (1c) or full NPs. Word-order is typically verb-initial in *bu-* and bare clauses, whilst the order is verb-medial in *ki-* clauses which potentially reflects their historical development via the reanalysis of cleft constructions.

In an initial pilot, we compared the usage of the three constructions in (1) in Kähler (1955) and a contemporary retelling of the same story:<sup>3</sup>

All finite clauses	<i>bu-</i>	bare	<i>ki-</i>	Total
Old Enggano	118 (c. 50%)	70 (c. 30%)	48 (c. 20%)	240
Contemporary Enggano	125 (c. 50%)	59 (c. 24 %)	63 (c. 26%)	247

<sup>3</sup> We excluded forms with embedded and non-finite *bu-* verbs without agreement from the count.

In many cases (e.g. following negation, imperatives, relative clauses) the choice of construction is syntactically constrained. If we exclude these from the counts, and focus only on main verbs where all of the constructions in (1) are possible, the patterns are as follows:

Main clauses	bu-	bare	ki-	Total
Old Enggano	106 (c. 58%)	55 (c. 30%)	21 (c. 12%)	182
Contemporary Enggano	118 (c. 62%)	42 (c. 22%)	31 (c. 16%)	191

In both Old Enggano and contemporary Enggano, *bu-* verbs are the most frequent in main clauses. These are often used at episode boundaries as the first clause in a new discourse unit and then followed by bare clauses:

- (2) ka-moho            e-paE            e'ana            ka-bu-pua,  
 3.NOM-different    DIR-child        DEM.MED        3.NOM-BU-run
- y-aha:E        y-aba-kiu            i-hoo            u-kãĩ            u-kepũũ  
 3.ERG-go        3.ERG-MOTION-hide    loc-inside        obl-base        obl-fern
- 'And the child fled, it went away and sought shelter on the inside of the foot of the fern.' (Kähler 1955).

This is particularly interesting as it demonstrates that the choice of *bu-* vs bare constructions may be determined by similar factors to symmetrical voice languages (cf. Riesberg et al. 2022), even though the mapping to subject does not change. Bare verbs are otherwise restricted to irrealis clause-types (questions, imperatives, negatives, hortatives) that reflect a possible reanalysis of voice morphology as marking TAM.

The major difference between Old Enggano and Contemporary Enggano is in the increase in the use of *ki-* clauses. Perhaps related, there is also an increase in the use of verb-medial order across all constructions. We argue that these changes most likely reflect the ongoing reanalysis of *ki-* from relativiser/subordinator to main clause marker, combined with the influence of contact with Indonesian varieties.

## References

- Brown, Lea. 2001. "A grammar of Nias Selatan." PhD Thesis, University of Sydney.
- Edwards, Owen. 2015. "The position of Enggano within Austronesian." *Oceanic Linguistics* 54 (1):54-109.
- Kähler, Hans. 1940. "Grammatischer abriß des Enggano." *Zeitschrift für Eingeborenen Sprachen* 30:81–117, 182–210, 296–320.
- Kähler, Hans. 1955. "Ein Text von der Insel Enggano (Westküste von Sumatra)." *Afrika und Übersee* 39:89–94.
- Kähler, Hans. 1957. "Ein Text von der Insel Enggano (Westküste von Sumatra)." *Afrika und Übersee* 41:153–56.
- Kähler, Hans. 1958. "Texte von der Insel Enggano (Westküste von Sumatra)." *Afrika und Übersee* 42:179-89.

- Kähler, Hans. 1960a. "Texte von der Insel Enggano (Westküste von Sumatra)." *Afrika und Übersee* 43:299–312.
- Kähler, Hans. 1960b. "Texte von der Insel Enggano (Westküste von Sumatra)." *Afrika und Übersee* 44:41–71.
- Kähler, Hans. 1961. "Texte von der Insel Enggano (Westküste von Sumatra)." *Afrika und Übersee* 45:205–24.
- Kähler, Hans. 1962. "Texte von der Insel Enggano (Westküste von Sumatra)." *Afrika und Übersee* 46:127–45.
- Kähler, Hans. 1964. "Texte von der Insel Enggano (Westküste von Sumatra)." *Afrika und Übersee* 48:46–57.
- Kähler, Hans. 1975. *Texte von der Insel Enggano (Berichte über eine untergehende Kultur)*. Berlin: Dietrich Reimer Verlag.
- Riesberg, Sonja. 2014. *Symmetrical voice and linking in western Austronesian languages, Pacific Linguistics*. Berlin: De Gruyter Mouton.
- Riesberg, Sonja, Maria Bardají i Farré, Kurt Malcher, and Nikolaus P. Himmelmann. 2022. "Predicting voice choice in symmetrical voice languages: All the things that do not work in Totoli." *Studies in Language*. 46 (2):453-516.

## (Avoiding) Overt reference to speech act participant objects: Evidence from Trans-Himalayan diachrony and spoken language corpora

Linda Konnerth  
Universität Bern

In TH, person indexation paradigms of (di)transitive verbs display a bewildering diversity of marking systems. In terms of alignment we find “hierarchical”-type patterns (i.e. displaying “co-argument sensitivity”, (Witzlack-Makarevich et al. 2016) as well as nominative-accusative-type patterns. However, the marking systems strikingly exhibit idiosyncratic alignment variants, certainly in the notoriously diffuse “hierarchical” but also in the nominative-accusative-type patterns (DeLancey 2017). The idiosyncrasy mostly revolves around the marking of Speech Act Participant (SAP) ‘objects’ (typically primary objects, treating equally transitive O and ditransitive R arguments). A convincing approach to make sense of the diversity and idiosyncrasies surrounding SAP object indexation on the verb points to the central role of sociopragmatic concerns as a catalyst for innovative paradigmatic verb forms (DeLancey 2018). Heath (1991; 1998), also along the lines of Brown and Levinson (1987), famously offers a negative motivation: irregular person indexation reflects the avoidance of direct reference to speaker (first person) and/or interlocutor (second person).

*Reference avoidance* is a useful concept to motivate the manifold creative constructions that can be seen springing up in person indexation paradigms across TH. At the same time, the opposite force to *overtly highlight* the involvement of speech act participants (even if only being affected participants in a propositionally broad sense) reliably surfaces up in the grammars of many languages of the world (Kuno 1987), including Trans-Himalayan. Both principles may thus be at play in diachronically shaping indexation paradigms (cf. Bickel et al. 1999). Furthermore, moving from the verb to arguments, we may also expect these principles to leave a language-specific synchronic footprint in the synchronic expression (‘highlighting’) or non-expression (‘avoidance’) of SAP object pronouns, based on the particular indexation pattern of that language.

In order to examine this hypothesis, I compare two languages with interestingly different systems of person indexation. Monsang (South-Central TH) exhibits a modern variant of the inherited Proto-South-Central/Proto-TH indexation system, with inverse marking in  $2 \rightarrow 1$  and  $3 \rightarrow \text{SAP}$  – in the latter configuration leaving the SAP reference actually ambiguous. In contrast, Karbi (unclassified TH) entirely lost the inherited person indexation system, but we find an intriguing innovative strategy of marking SAP objects (and occasionally SAP’s in some other non-subject participant roles), but nothing else, on the verb. In the  $3 \rightarrow \text{SAP}$  scenarios, both languages contain SAP object marking on the verb but in an ambiguous fashion that does not specify which SAP the object is. While this is the same in both languages, the indexation systems are otherwise strikingly different. As such, the respective paradigmatic morphological contexts suggest that we find Monsang associated with *avoidance* but Karbi with *highlighting* of SAP objects. In the context of these two different indexation systems, I present a qualitative analysis of the expression of SAP object pronouns in corpora of spoken Monsang and spoken Karbi. Interestingly, we see that the synchronic corpus data paint a picture that resonates with the morphological-historical analysis of the verbal indexation paradigms.

## References

- Bickel, Balthasar, Walter Bisang, and Yogendra P. Yādava. 1999. "Face vs. Empathy: The Social Foundation of Maithili Verb Agreement." *Linguistics* 37 (3): 481–518.
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- DeLancey, Scott. 2017. "Hierarchical and Accusative Alignment of Verbal Person Marking in Trans-Himalayan." *Journal of South Asian Languages and Linguistics* 4 (1): 85–105.
- . 2018. "Deictic and Sociopragmatic Effects in Tibeto-Burman SAP Indexation." In *Typological Hierarchies in Synchrony and Diachrony*, edited by Sonia Cristofaro and Fernando Zúñiga, 343–76. Amsterdam/Philadelphia.
- Heath, Jeffrey. 1991. "Pragmatic Disguise in Pronominal-Affix Paradigms." In *Paradigms: The Economy of Inflection*, edited by Frans Plank, 75–89. Berlin/New York: Mouton de Gruyter.
- . 1998. "Pragmatic Skewing in 1↔2 Pronominal Combinations in Native American Languages." *International Journal of American Linguistics* 64 (2): 83–104.
- Kuno, Susumu. 1987. *Functional Syntax: Anaphora, Discourse and Empathy*. Chicago: University of Chicago Press.
- Witzlack-Makarevich, Alena, Taras Zakharko, Lennart Bierkandt, Fernando Zúñiga, and Balthasar Bickel. 2016. "Decomposing Hierarchical Alignment: Co-Arguments as Conditions on Alignment." *Linguistics* 54 (3): 531–61.

## Voice choice in Äiwoo discourse

Åshild Næss  
University of Oslo

The Oceanic language Äiwoo has a symmetrical voice system with two basic voices, an actor voice and an undergoer voice; in other words, with most transitive verbs, the speaker has a choice to make either the actor or the undergoer argument the ‘subject’ (for a discussion of the suitability of the term ‘subject’ in Äiwoo and other symmetrical voice languages, see Schachter 1976, 1977, Riesberg 2014, Næss 2015). In actual discourse, as represented by a corpus of around 9 hours of spoken language collected by the author, combined with written materials collected in the 1970s and 80s by Stephen Wurm and archived in PARADISEC ([paradisec.org.au](http://paradisec.org.au)), the use of the two constructions is, however, significantly skewed: the actor voice is used only around 14% of the time. While it is common in Austronesian symmetrical voice languages for the undergoer voice to be more frequent than the actor voice in usage, this number is unusually low.

Given these figures, it seems reasonable to treat the undergoer voice as the default transitive construction, and to try to account for the use of the actor voice: under what circumstances do Äiwoo speakers choose the actor rather than the undergoer argument as the subject?

A range of factors appear to be relevant, and these overlap to a large extent with what has been found in other symmetrical voice languages: the actor voice is typically required when the actor argument is focused or relativized on (though unlike in many other symmetrical voice languages, the latter is not without exceptions), and is preferred in various contexts where reference is primarily to the action as such rather than its effect on an undergoer, such as with action nominalisation, various types of manner expressions, and when there is a change of activity in a sequence of actions; the latter may have something in common with the use of the actor voice to indicate a new episode in discourse in the symmetrical voice language Totoli (Riesberg et al. 2022). However, the clearly most frequent context is when the undergoer argument is indefinite or nonreferential, in other words, when the actor argument is more referentially prominent than the undergoer. This points towards the possible reanalysis of the actor voice as an intransitive construction, since argument referentiality is a key aspect of transitivity as it is defined in the typological literature (Hopper and Thompson 1980, Kemmer 1993, Næss 2007). This is a reanalysis that we know has happened in most Oceanic languages, where the original undergoer voice has become the default transitive construction, and the actor voice has been reanalysed as intransitive. The usage patterns in present day Äiwoo can thus point to a possible path by which such a reanalysis may have taken place.

### References

- Hopper, Paul J., and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56:2, 251-299.
- Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam: John Benjamins.
- Næss, Åshild. 2007. *Prototypical transitivity*. Amsterdam: John Benjamins.
- Næss, Åshild. 2015. Voice at the crossroads: symmetrical clause alternations in Äiwoo, Reef Islands, Solomon Islands. *Oceanic Linguistics* 54:1, 270–307.

- Riesberg, Sonja. 2014. *Symmetrical voice and linking in Western Austronesian languages*. Boston/Berlin: de Gruyter Mouton.
- Riesberg, Sonja, Maria Bardají i Farré, Kurt Malcher and Nikolaus P. Himmelmann. 2022. Predicting voice choice in symmetrical voice languages: all the things that do not work in Totoli. *Studies in Language* 46(2), 453–516.
- Schachter, Paul. 1976. The subject in Philippine language is: topic, actor, actor topic, or none of the above. In Charles Li (ed.), *Subject and topic*, 491-518. New York: Academic Press.
- Schachter, Paul. 1977. Reference-related and role-related properties of subjects. In Peter Cole and Jerrold M. Saddock (eds.), *Syntax and Semantics. Volume 8: Grammatical relations*, 279-306. New York: Academic Press.

# Valency-encoding devices in a spoken Northeastern Neo-Aramaic corpus

Maria Ovsjannikova and Sergey Say  
University of Potsdam

**Framework.** The main function of valency-encoding devices, such as cases, adpositions and verbal indices, is to establish links between referents in discourse and their argument roles. However, most approaches to valency classes are focused on verbal lexica, where individual verbs are viewed as distinct **types** associated with specific patterns (Malchukov & Comrie 2015). By contrast, there is a clear lack of **token-based** studies focused on the use of valency-encoding devices in discourse, where the actual transmission of information takes place. Besides, most typologically-oriented studies of valency in discourse focus upon major clause types, such as one-argument intransitive clauses and transitive clauses, as is the case of the so-called Preferred Argument Structure hypothesis (DuBois 1987), or its criticism (Haig, Schnell 2016). In our study, we aim at partially filling these gaps.

**Goals.** In our study, we explore the use of argument-encoding devices in Northeastern Neo-Aramaic (NENA) varieties spoken by “Assyrians” in Russia, highlighting differences between canonical core arguments (A, S, O) and oblique/non-canonical arguments in terms of their preferred discourse niches.

**Language.** NENA varieties have flexible (mainly SOV and SVO) word order and lack morphological case. A and S arguments are always indexed and never flagged. O arguments simultaneously display differential object flagging and differential object indexing. Non-core arguments are not indexed and typically are flagged by prepositions.

**Data and methods.** The data come from a smallish corpus of semi-spontaneous oral texts, mostly narratives, recorded in “Assyrian” settlements in the South of Russia (Ovsjannikova et al. 2022-). Clauses extracted from the corpus were additionally annotated for the use of argument-encoding devices and for the type of referential expressions. The annotation system was largely inspired by GRAID (Haig et al. 2021).

**Results and hypotheses.** The results obtained so far are preliminary, as they are based on the work in progress. In accordance with expectations, the choice between full NPs and referential zeros (pronouns are rarely used in NENA discourse) is not evenly distributed among arguments. O and non-core arguments favour full NPs if compared to S and A arguments. However, in some respects, non-core arguments occupy a discourse niche that differentiates them from O arguments. For example, the use of overt non-core arguments, typically accompanied by a preposition, correlates negatively with the use of full NPs in core positions, most notably in the S slot. Arguably, this pattern can be seen as a trade-off scenario: conveying more information in one domain makes it possible to convey less information in the other domains (Koplenig et al. 2017). In our talk, we are going to discuss some other phenomena observed in the NENA discourse in a similar vein. In particular, we will discuss patterns in which non-core arguments are not overtly flagged and will speculate that this is possible when the relevant information is predictable, that is, retrievable from either the meaning of the predicate (possessive constructions) or the noun phrase itself (e.g. in the case of some locative nouns).



## References

- Du Bois, J. W. 1987. The discourse basis for ergativity. *Language* 63(4), 805–855.
- Haig, G. & S. Schnell. 2016. The discourse basis of ergativity revisited. *Language* 92(3). 591–618.
- Haig, G., S. Schnell, & N. N. Schiborr. 2021. Universals of reference in discourse and grammar: Evidence from the Multi-CAST collection of spoken corpora. In: G. Haig, S. Schnell, & F. Seifart (eds.). *Doing corpus-based typology with spoken language corpora: State of the art*. Honolulu: University of Hawai'i Press. 141–177.
- Haspelmath, M. 2015. Transitivity prominence. In: A. L. Malchukov & B. Comrie (eds.), *Valency classes in the world's languages, vol. 1: Introducing the framework, and case studies from Africa and Eurasia*. Berlin: de Gruyter. 131–147.
- Koplenig, A., P. Meyer, S. Wolfer & C. Müller-Spitzer. 2017. The statistical trade-off between word order and word structure – Large-scale evidence for the principle of least effort. *Plos One* 2017(12).
- Ovsjannikova, Maria, Kirill Kozhanov, Natalia Logvinova, Alexey Lyavdansky, Alina Russkix, Sergey Say, Varvara Shuvalova, Elena Shvedova, and Elizaveta Zabelina. 2022–. *Corpus of NENA varieties spoken in Russia*. Available online at <https://nena-dict.ru/corpus>.

## DCM in V-final languages as local communicative instructions: ergative(?) in Anal Naga

Pavel Ozerov

University of Innsbruck

Verb-final Trans-Himalayan languages are well-known for adnominal markers, whose distribution is driven by elusive pragmatic factors. This is the case with differential case markers, which are commonly related to multiple competing semantic and pragmatic phenomena (e.g. Chelliah and Hyslop 2011). This paper presents a case study from Anal Naga (South-Central; Manipur) that links such marking – and the elusiveness of its functions – to interaction-managing and production-oriented factors. The data comes from a multimedia corpus of everyday interaction. The clitic =tỹ: is a differential ergative marker in the language. It is commonly obtained in elicitation as a case-marker of agentive arguments, while the marking of other arguments by =tỹ: is rejected as unacceptable (including agentive S). Its occurrence appears to be related to such vague interpretations as emphasised agency, new topics, contrast, and overall importance of the referent.

In the proposed analysis, the function of =tỹ: lies in the domain of local interaction-managing instructions: when the current interactional move is restricted to merely naming the referent, the marker is used to preliminarily foreshadow its agentive role in the follow up discourse. In other words, the marking signals the upcoming agentive role (in the sense of causing effect on other entities) of the named referent, when the NP expressing this referent occupies a separate Intonation Unit (IU). In such cases, the interactional move performed by this prosodically detached NP can often be observed through converging interactional cues. It is also commonly evident that the rest of the sentence structure is not yet planned at this stage. Thus, rather than representing case in its traditional view, =tỹ: functions as a local cue guiding the interpretation of the NP in the unfolding structure and foreshadowing the referent's agentive role in the preliminary conceptualisation of the event. Since agentivity plays central role in communication (Bickel et al. 2015), but cannot be interpreted from the NP position due to the V-final syntax of the language, it is signalled by an explicit marker.

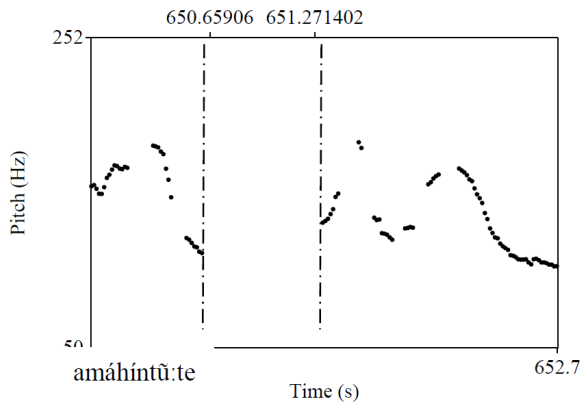
In the examined data, =tỹ: is found overwhelmingly (80%) at the end of IUs which perform a local interactional move, as is evident from such factors as following pauses, hesitations, negotiation of reference, gaze shifts and gesture. This is the case in (1), where the reference to a given participant is followed by a hesitation pause. Notably, the example also shows that the marker does not restrict the follow-up construal to a transitive verb (as a clause-level syntactic marker would be expected to do). Instead, the agentivity-foreshadowing effect of =tỹ: can last into later, non-immediate talk. On the contrary, agentive NPs unmarked by =tỹ: are overwhelmingly produced within a single intonation unit (IU) with the verb (80% of such cases; in the remaining 20% the NP is marked by other pragmatic particles, primarily contrastive or additive). This is illustrated in (2), where the brand new but unmarked A-argument 'Samuel' forms an accental phrase within a smoothly produced IU.

The apparent pragmatic functions associated with the marker are epiphenomenal of its interaction managing function. For instance, NPs are commonly expressed in separate IUs if the

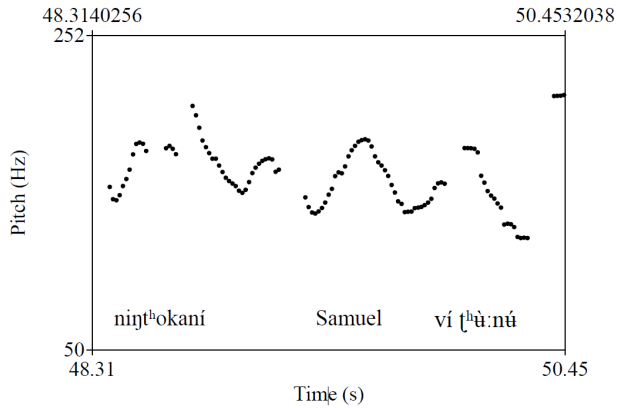
speaker aims at centering joint attention at them. Consequently, =tỹ: is commonly found with expressions of referents that are not yet at the centre of joint attention (i.e. non-given). Yet, 50% of agentive NPs unmarked by =tỹ: express nonetheless new referents (as in (2)), while expressions of given referents can be marked by =tỹ: if their separation is triggered by other factors (e.g. planning troubles, as in (1)). Similarly, NP separation can be a result of substantial amount of competing information related to the named referent (as in (1)), thus epiphenomenally rendering the effect of discourse salience. Hence, this study demonstrates how tackling pragmatic adnominal markers from a usage-based interactionally-informed perspective, which considers prosodic, production-related, multimodal, and interaction-managing factors, can account for their elusive pragmatic-semantic properties.

Examples (all from anm\_20161013\_Jm\_pastors1)

- (1) ‘As we were talking over there’  
 amá-hín=**tỹ**:=te... va-at<sup>h</sup>á-há háŋ-kàl-nana-jè-né jũ-he  
 3-PL=**ERG**=CNTR 3-upper.part-DIR up-climb-INSTR-PL-N.FUT wine-PROX1  
 va-né:li-va-je=hebo  
 3-drink-INADV-N.FUT-3PL=DM  
 ‘They=**tỹ**: (...612ms pause...) climbed to the upper floor, drank the wine...’



(2) niŋtʰokaní **Samuel** ví tʰù:-nú, sá i-dò:n  
 Monday PN dog take.along-N.FUT animal NMLZ-hunt  
 i-vá-dó:-va  
 NMLZ-go-away-COP  
 ‘On Monday Samuel took the dogs and went hunting.’



## References

- Bickel, Balthasar, Alena Witzlack-Makarevich, Kamal K. Choudhary, Matthias Schlesewsky, and Ina Bornkessel-Schlesewsky. 2015. “The Neurophysiology of Language Processing Shapes the Evolution of Grammar: Evidence from Case Marking.” *PLOS ONE* 10 (8): e0132819. <https://doi.org/10.1371/journal.pone.0132819>.
- Chelliah, Shobhana L., and Gwendolyn Hyslop. 2011. “Introduction to Special Issue on Optional Case Marking in Tibeto-Burman.” *Linguistics of the Tibeto-Burman Area* 34 (2): 1–7.

## Post-verbal subjects in an SV language: non-agentivity and discourse structure

Linn Iren Sjønes Rødvand  
University of Oslo

**Summary:** This paper explores post-verbal subject expressions in Patani (SHWNG, Austronesian), a surprising phenomenon in an otherwise SV language. While post-verbal subjects are restricted to non-Actor verbs and these occurrences as such are reminiscent of semantic alignment systems, the synchronic use of post-verbal subjects also seems to be connected to discourse structure and grounding.

The Patani language is spoken on the island of Halmahera, North-Maluku, Indonesia. So far, this language has been poorly documented, but this paper is based on data (mostly spontaneous speech) collected by the author in 2019 and 2021. At first sight, Patani has a straightforward nominative-accusative alignment, which is the most widespread alignment pattern in SHWNG (Gasser et al. in press): The dominant word order in Patani is SV in intransitive clauses, and AVP in transitive clauses, and both the S argument and the A argument are indexed on the verb through a prefix. However, there are exceptions to this main pattern. Most notably, a pronominal form of the S argument occasionally follows the verb ( $VS_{PRO}$ ). This may come in addition to the subject prefix (lexically and pragmatically conditioned) and a preposed subject NP (pragmatically conditioned). As a result, the subject of an intransitive verbal predicate can be coded in three different ways, partly depending on verb type.

Prefix	Prefix + $VS_{PRO}$	$VS_{PRO}$
1) <i>n-cep</i> 3SG-bathe 'He bathes'	2) <i>n-pólón i</i> 3SG-return 3SG 'He returns'	3) <i>tér-pung i</i> IA-break 3SG 'It breaks'
<i>Found with:</i> Most verbs	<i>Found with:</i> Non-Actor verbs Intradirective verbs	<i>Found with:</i> Verbs with the prefix <i>tér-</i> (involuntary action)

Table 1: the three coding strategies of S in Patani.

The system sketched out in Table 1 above is reminiscent of patterns found in languages which are described as having split-S/semantic alignment in the region (see Klamer 2008, Holton 2008, Schapper 2015), where the encoding of the S argument depends on the lexical aspect of the verb (e.g. dynamic vs. non-dynamic) and/or agentive/patientive characteristics of the participant. In these languages, the encoding of the S argument is either identical to that of a P argument (like example 3 above), or it has double encoding as both A and P (like 2 above). The Patani system diverges from semantic alignment systems in that non-Actor verbs and intradirective verbs only occasionally occur with coding strategy 2; most commonly strategy 1 is chosen. Preliminary data suggests that the choice of coding strategy 2 is influenced by 1) lexical factors: for instance, *pólón* 'return' is much more likely to occur with coding strategy 1 than *wé* 'return'; 2) pragmatic factors: strategy 2 is marked in terms of frequency and can therefore be used to express additional,

pragmatic meaning. For instance, *n-fan i* '3sg-go 3sg' can have the meaning 'he escaped' rather than just 'he went'; 3) discourse structural factors: coding strategy 2 may be used to signal the end of a discourse-structural unit which furthermore serves a backgrounding function.

These hypotheses will be further tested during fieldwork *primo* 2023, primarily through collection and analysis of narratives prompted by non-linguistic stimuli targeting intradirective and non-Actor verbs. As Riesberg et al. (2022: 509) note, discourse-structural units need to be defined on independent grounds. Arriving at an operationalizable definition of the relevant discourse units will therefore be an important goal in the analysis of the new narrative data.

## References

- Gasser, Emily, Laura Arnold, and David Kamholz. In press. The languages of Halmahera, Raja Ampat, and west New Guinea. To appear in Alexander Adelaar, Antoinette Schapper (Eds.), *The Oxford Guide to the Malayo-Polynesian languages of Southeast Asia*. Oxford: Oxford University Press.
- Holton, Gary. 2008. The rise and fall of semantic alignment in North Halmahera, Indonesia. In Mark Donohue, Søren Wichmann (Eds). *The Typology of Semantic Alignment*. 252–276. Oxford: Oxford University Press.
- Klamer, Marian. 2008. The Semantics of Semantic Alignment in eastern Indonesia. In Mark Donohue, Søren Wichmann (Eds), *The Typology of Semantic Alignment*. 221–251. Oxford: Oxford University Press.
- Schapper, Antoinette. 2015. Wallacea, a Linguistic Area. *Archipel* 90(2), 99–151.
- Riesberg, Sonja, Maria Bardají i Farré, Kurt Malcher, and Nikolaus P. Himmelmann. 2022. Predicting voice choice in symmetrical voice languages. All the things that do not work in Totoli. *Studies in Language* 46(2), 453–516.

## Non-standard features in spoken Germanic language varieties – the cases of null subjects and negative concord

Henrik Rosenkvist  
University of Gothenburg

The Germanic languages probably constitute the one language group that has been described in greatest detail – they have a long literary history and have been studied for centuries. Still, the descriptions of these languages are in general based on the standardized written forms, and for this reason some linguistic features that are relatively widespread in various vernaculars have remained quite undetected until recently. In this talk, I will focus on two grammatical constructions that have been argued not to exist in Germanic and north Germanic, respectively: null subjects and negative concord.

It has been claimed that the V2-parameter, that characterize the Germanic languages (but not English), is incompatible with null subjects (cf. Jaeggli & Safir 1989:33, Rohrbacher 1999:251ff), a statement which rests on the observation that no standard Germanic language allows thematic null subjects. While this is true, it is also true that null subjects do occur in a number of modern Germanic mainly non-written vernaculars, such as for instance Övdalian, Swabian and Frisian. Furthermore, the distribution of null subjects in these language varieties corresponds with verb morphology in a similar fashion across the board. It is argued that these null subjects should be considered innovations, and not continuations of null subjects in Old Germanic, due to some grammatical differences as well as diachronic data (but cf. Walkden 2014 etc.). A common development in the Old Germanic languages is namely that null subjects disappeared one or two centuries after the first written records, and this may possibly be ascribed to the process of *Verschriftligung* (Koch & Oesterreicher 1985).

Another feature that is prevalent in Germanic vernaculars but missing from all standard languages is negative concord (NC). While the presence of NC in previous stages of West Germanic has been widely discussed, NC in North Germanic has not been acknowledged: “There is no cumulative or multiple negation, either in standard speech or in the dialects [...]” (Haugen 1986:157). However, NC can be attested in vernaculars such as Övdalian, Ostrobothnian and Estonian Swedish. NC was a regular feature of most West Germanic standard languages until modern times, when it was ousted due to prescriptive standardization (Hoeksema 1997, Langer 2001, Ingham 2006, Trudgill 2009; but cf. Zeiljstra 2016).

Accordingly, there may be different explanations for the non-existence of null subjects and NC, respectively, in the standard languages. However, from a more general perspective, it seems that both processes mainly target grammatical structures where constituents are omitted or doubled (cf. Brandner 2012) – but, crucially, this only occur in written languages. Hence one may compare written and standardized languages with well-kept gardens – they may be nice and orderly, but some plants are considered weeds and are systematically rooted out; such plants only thrive in natural environments.

## Do difficulties of reference production differ across grammatical relations? An examination of the distribution of hesitation pauses in eight diverse languages.

Stefan Schnell<sup>1</sup>, Frank Seifart<sup>2,3</sup>, Nils N. Schiborr<sup>4</sup>, Geoff Haig<sup>4</sup>  
University of Zurich<sup>1</sup>, CNRS<sup>2</sup>, Leibnizzentrum Allgemeine Sprachwissenschaft<sup>3</sup>, University of Bamberg<sup>4</sup>

A long tradition of functionalist research in discourse and grammar has determined certain correlations between grammatical relations and reference (Chafe 1976; Givón 1976; Dixon 1979; Cooreman, Fox, Givón 1984) so that subjects {S,A} are associated with topic continuity (Givón 1983; Shibatani 1991) and {S,P} with ‘focality’ (DuBois 1987, 2003) or – for deep ergative languages – topic continuity (Dixon 1979). The function A alone has been found to preferably host continuing topics (Cooreman, Fox, Givón 1984; DuBois 1987; Haig & Schnell 2016), and P to be a likely host for new referents (Schnell, Schiborr, Haig 2021). The ‘pragmatic linking’ (Durie 2003) of grammatical relations in syntax is also reflected in models of reference processing like Centering Theory (Grosz, Joshi, Weinstein 1995) and has been regarded as a central motivation for the evolution of grammatical relations (Givón 1979; Shibatani 1991; cf. also Evans & Levinson 2009:440).

Such correlations are postulated to reflect ‘soft constraints’ and hence mere preferences on language production; the dispreference of deviating constellations are explained in terms of higher processing costs that producers actively seek to avoid. For instance, a full lexical noun phrase with a new referent in A function is on this view avoided due to the processing challenges it will pose to the comprehender (DuBois 2003).

Despite its tremendous impact on our understanding of grammatical relations, functionalist work on discourse and grammar has notoriously failed to establish evidence for these postulated processing difficulties. In this contribution we take up this challenge by examining the distribution of hesitation pauses which have been regarded as indicators of planning difficulties for the producer (Fox Tree & Clark 1997, de Jong 2016). Specifically, hesitation pauses in noun production have been argued to be related to paradigmatic complexity and possibly referential choice (Seifart et al 2018).

Here, we zoom in on the question of referential choice by considering differences in referential status (new vs. old) and postulated differences in processing efforts across syntactic functions, focussing here on A vs. P: as new referents are assumed to be harder to process we expect hesitation pauses to be generally more likely before NPs with new than with old reference. And as the A function is generally associated with high accessibility, we expect a lowaccessibility NP in A to be more likely to incur a hesitation pause due to planning difficulties than a low-accessibility NP in P function.

We investigate spoken discourse production data from eight different languages which have been coded for morphosyntactic structure (Haig & Schnell 2014) and reference (Schiborr, Schnell, Thiele 2018) and whose textual representation is time-aligned on phonemic segment level (Paschen et al. 2020): Bora (Seifart 2022), English (Schiborr 2022), Nafsan (South Efate) (Thieberger 2022), Northern Kurdish (Kurmanji) (Haig, Vollmer & Thiele 2022), Sanzhi Dargwa (Forker & Schiborr 2022), Tabasaran (Bogomolova, Ganenkov & Schiborr 2022), Teop (Mosel



2022), Vera'a (Schnell 2022), Yali (Riesberg 2022). This annotation enables exact measurements of hesitation pauses in relation to different nominal expressions across the two functions A and P.

Figures 1-2 present preliminary results regarding hesitation pauses preceding lexical NPs across different information statuses (Figure 1) and syntactic functions (Figure 2), excluding main-clause initial NPs, since pauses before these are likely to reflect planning of larger units.

These preliminary findings display considerable degrees of cross-linguistic variation regarding the distribution of hesitation pauses, which is what we expect given that the corresponding grammatical patterns are 'soft constraints'. Despite this variation, there are also tendencies detectable. Specifically, new information status tends to be associated with higher proportions of pauses (Figure 1), as expected. On the other hand, lexical NPs in A function seem to be associated with more hesitation pauses, reflecting more planning effort than those in P function (Figure 2). While lending some tentative support to pragmatic linking as described above, these findings also raises the question how planning difficulties on behalf of producers are to be interpreted in this context as these may not necessarily result from considerations of audience design, and we will elaborate on this point in our contribution.

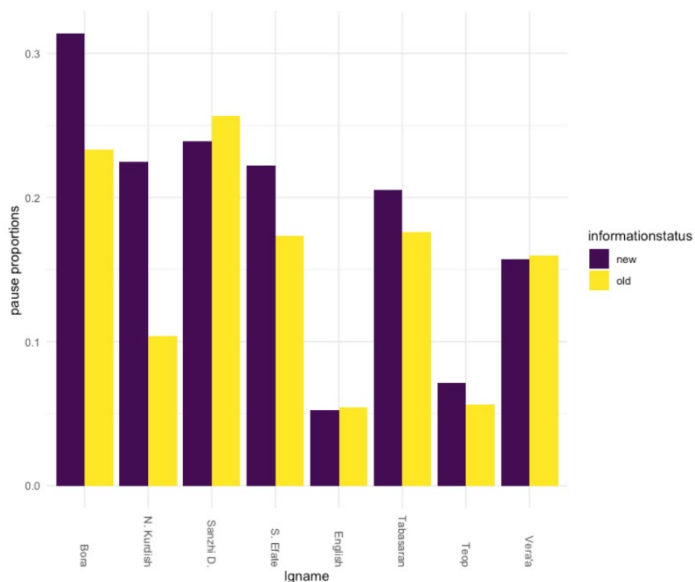


Figure 1: Hesitation pauses preceding lexical NPs with different referential statuses ('new' means either brand-new or bridging contexts).

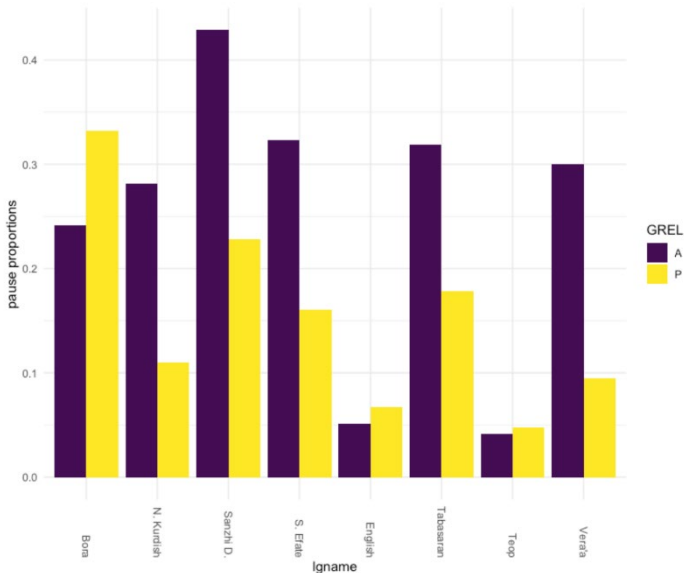


Figure 2: Hesitation pauses preceding lexical NPs with different grammatical relations

## References

- Ariel, M. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Bogomolova, N., Ganenkov, D., Schiborr, N.N. 2022. Tabasaran DoReCo Dataset. In F. Seifart, L. Paschen, L., M. Stave (Eds.), *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).
- Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic*. 25–55. New York: Academic Press.
- Cooreman, A., Fox, B.A., Givón, T. 1984. The discourse definition of ergativity. *Studies in Language*. 8 (1): 1–34.
- Dixon, R. M. W. 1979. Ergativity. *Language* 55(1). 59–138.
- Du Bois, J. W. 1987. The discourse basis of ergativity. *Language* 805–855.
- Du Bois, J. W. 2003. Discourse and grammar. In M. Tomasello (Ed.), *The new psychology of language, vol. 2*, 47–88. Mahwah, NJ: Erlbaum.
- Durie, M. 2003. New light on information pressure: Information conduits, 'escape valves', and role alignment stretching. In J. W. Du Bois, L. Kumpf, W. J. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function*. 159–196. Amsterdam: John Benjamins.
- Evans, N., Levinson, S. C. 2009. The myth of language universals. *Behavioral and Brain Sciences* 32(5). 429–492.
- Forker, D., Schiborr, N.N. 2022. Sanzhi Dargwa DoReCo Dataset. In F. Seifart, L. Paschen, M. Stave (Eds.), *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).
- Fox Tree, J.E., Clark, H.H. 1997. Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition* 62 (2), 151–167.

- Givón, T. 1976. Topic, pronoun and grammatical agreement. In C. N. Li (Ed.), *Subject and Topic*, 149–188. London: Academic Press.
- Givón, T. 1979. From discourse to syntax: grammar as a processing strategy. In T. Givón (Ed.), *Discourse and syntax*, 81–112. Brill.
- Givón, T. 1983. Topic continuity in discourse. In T. Givón (Ed.), *Topic continuity in discourse, vol. 3*, 1–42. Amsterdam: John Benjamins.
- Grosz, B.J., Joshi, A.K., Weinstein, S. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21 (2): 203–25.
- Haig, G., Schnell, S. 2014. Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators (version 7.0) <https://multicast.aspra.uni-bamberg.de/#annotations>.
- Haig, G., Schnell, S. 2016. The discourse basis of ergativity revisited. *Language* 92(3). 591–618.
- Jong, N.H. de. 2016. Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency." *International Review of Applied Linguistics in Language Teaching* 54 (2): 113–132.
- Haig, G., Vollmer, M., Thiele, H. 2022. Northern Kurdish (Kurmanji) DoReCo Dataset. In F. Seifart, L. Paschen, M. Stave (Eds.), *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).
- Mosel, U. 2022. Teop DoReCo Dataset. In F. Seifart, L. Paschen, M. Stave (Eds.), *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).
- Paschen, L., Delafontaine, F., Draxler, C., Fuchs, S., Stave, M., Seifart, F. 2020. Building a time-aligned cross-linguistic reference corpus from language documentation data (DoReCo). *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC'20)*, Marseille, France, 13–16 May 2020, 2657–2666.
- Riesberg, S. 2022. Yali (Apahapsili) DoReCo Dataset. In F. Seifart, L. Paschen, M. Stave (Eds.), *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).
- Schiborr, N. N. 2022. English (Southern England) DoReCo Dataset. In F. Seifart, L. Paschen, M. Stave (Eds.), *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).
- Schiborr, N. N., Schnell, S., Thiele, H. 2018. RefIND — Referent Indexing in Natural-language Discourse: Annotation guidelines (v1.1).", University of Bamberg.
- Schnell, S. 2022. Vera'a DoReCo Dataset. In F. Seifart, L. Paschen, M. Stave (Eds.), *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).
- Schnell, S., Schiborr, N. N., Haig, G. 2021. Efficiency in discourse processing. *Linguistics Vanguard* 7(s3).

- Seifart, F. 2022. Bora DoReCo Dataset." In F. Seifart, L. Paschen, M. Stave (Eds.), *Language Documentation Reference Corpus (DoReCo) 1.2*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).
- Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., Witzlack-Makarevich, A., Jong N. H. de, Bickel, B. 2018. Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences* 115(22). 5720–5725.
- Shibatani, M. 1991. Grammaticization of topic into subject." In E. C. Traugott, B. Heine (Eds.), *Approaches to Grammaticalization*. 93–133. Amsterdam; Philadelphia: John Benjamins.

# Voice type and argument realization of applicatives in Javanese discourse

Jozina Vander Klok  
Humboldt-Universität zu Berlin

**Introduction.** This study investigates applicatives within discourse in Javanese (Austronesian; Malayo-Polynesian; Western Indonesian), which given the multiple strategies for expressing the applied argument, present a window into understanding the factors that motivate how speakers use these strategies across voice type. The **applied argument** in Javanese can be within the verb phrase as the direct object (as per the prototypical function of applicativization) or headed by a prepositional phrase (PP) in Actor Voice (AV; see (1a-b)), or as the subject in an Undergoer Voice (UV), indicated by *di-* in (2) (e.g., Sofwan 2010).

- (1) a. Ali n-jupuk-ake **Aminah** buku.  
Ali AV-take-APPL Aminah book  
'Ali took a book for Aminah.'
- b. Ali n-jupuk-ake buku **kanggo Aminah**.  
Ali AV-take-APPL book for Aminah  
'Ali took a book for Aminah.' (Sofwan 2010: 58)
- (2) **Petrus** di-jupuk-ake buku dening Jana  
Petrus UV-take-APPL book by Jana  
'The book was taken for Petrus by Jana' (Sofwan 2010: 59)

**Research questions. [1]** What are the factors affecting voice choice of applicatives in Javanese discourse? Previous work on narrative texts in Madurese (Davies 2005) and *Tukang Besi* (Donohue 2001) suggests that the primary role of applicatives is to make it possible for the applied argument—typically a semantic peripheral role like beneficiary, location, or goal—to be a subject; thereby interacting with voice (cf. (2) for Javanese).

**[2]** What are the factors affecting overt arguments in applicatives across voice type? Previous work on Cirebon Javanese, shows that null arguments are the norm in transitives (Ewing 2005). Overt subjects are used for information-structural reasons (to establish a topic, present a contrast (focus/topic)), stance (express an opinion), and reporting speech (Ewing 2014). The reasons for an overt object has not been investigated.

**Methodology.** The dataset is based on a compiled subcorpus of conversational data of Javanese discourse: 3 conversations recorded in Semarang in 2016-2017, and 2 conversations from the Jakarta Field Station Language Archive; one from Semarang in 2014 and one from Pematang in 2004. These are chosen as within the Central Javanese dialect grouping and because they are all of the same genre as conversation, in order to relate to previous work on Cirebon Javanese conversational data (Ewing 2005, 2014).

Javanese has two types of applicatives, locative *-(n)i* and benefactive *(a)ke/no*. Because Indonesian is also frequently used in present-day Javanese discourse (cf. Errington 1998; Goebel

2005), Indonesian applicative counterparts (*-(n)i* and *-kan*) were also included. The total number of tokens in this corpus is 102 locatives (85 JAV, 17 IND), and 103 benefactives (65 JAV, 38 IND). **Results [1].** The most striking factor on the usage of voice type across applicatives in Javanese is language (Table 1). Javanese applicatives have higher use of UV, wherein the applied argument is the subject, compared AV, as shown with 59% for locatives and 64% for benefactives. This result is similar to in related languages (cf. Davies 2005 for Madurese; Donohue 2001 for *Tukang Besi*). The use of an Indonesian applicative within this discourse (given the multilingual context), however, shows the opposite pattern, whereby AV is used the most (82% of locatives and 69% of benefactives). This finding may relate to the higher prevalence of AV in Indonesian (e.g., Shiohara 2015; Djenar 2018).

Applicative	Language	AV	UV
"Locative"	Javanese -I	41%	<b>59%</b>
"Benefactive"	Javanese -KE/-NA	36%	<b>64%</b>
"Locative"	Indonesian -I	<b>82%</b>	18%
"Benefactive"	Indonesian -KAN	<b>69%</b>	31%

**Results [2].** Preliminary results based on 3 conversations (44 tokens for *-i*, 46 token for *-ake/-kan*) concerning argument realization of applicatives in Javanese discourse support previous findings that null arguments are the norm in Javanese discourse, see Table 2 (cf. Ewing 2005, 2014; Malihah 2018). Most often overt subjects with applicatives seem to (re-)establish a topic, in line with findings from Donohue (2001). Other functions such as present a contrast (focus/topic), stance, or reported speech, are not represented in these initial results. Additional factors including animacy and semantic role, as well as in relation to the direct object and PP, will be investigated in developing these results further. Despite the fact that UV is more often used for applicatives in Javanese discourse, the applied argument is not more likely to then be overtly expressed.

Subject	JAV/IND <i>-i</i>		JAV <i>-ake</i> / IND <i>-kan</i>	
	AV	UV	AV	UV
<i>overt</i>	8%	2%	16%	15%
<i>null</i>	92%	98%	84%	85%

## References

- Davies, William. 2005. The richness of Madurese voice. In I Wayan Arka and Malcolm Ross, (eds.), *The many faces of Austronesian voice systems: some new empirical studies*, 197–220. Canberra: Pacific Linguistics.
- Donohue, Mark. 2001. Coding choices in argument structure: Austronesian applicatives in texts. *Studies in Language* 25:2, 217–254.
- Errington, J. Joseph. 1998. *Shifting languages: Interaction and identity in Javanese Indonesia*. Cambridge: Cambridge University Press.
- Ewing, Michael C., 2005. *Grammar and Inference in Conversation: Identifying Clause Structure in Spoken Javanese*. Amsterdam: John Benjamins.

- Ewing, Michael C., 2014. Motivations for first and second person subject expression and ellipsis in Javanese conversation. *Journal of Pragmatics* 63. 48-62.
- Goebel, Zane. 2005. An ethnographic study of code choice in two neighbourhoods of Indonesia. *Australian Journal of Linguistics* 25. 85–107.
- Malihah, Noor. 2018. The description of the *di-* passive construction in dialectal Javanese. *Wacana* 19. 122-148.

## Factors in differential subject marking in Uruangnirin: functional, structural and social

Eline Visser

University of Oslo and Lund University

Uruangnirin is an Austronesian language spoken in eastern Indonesia. The subject is marked in one of the following ways: by means of a pronoun only, by means of a pronoun and a subject prefix, or by a subject prefix only. For most person/numbers, there are at least two subject prefixes. In this talk, I will present the factors that seem to play a role in this differential subject marking. (1) illustrates the phenomenon with four different ways to mark the first person subject with the verb *fasa* 'to buy'.

- (1) a. *meing lau lei-fasa*  
oil 1SG 1SG-buy  
'The oil I buy.'
- b. *kalo lei-fasa wai am kilo ni harga beda*  
if 1SG-buy stuff LOC Kilo 3SG.POSS price different  
'If I buy stuff in Kilo the price is different.'
- c. *peta li-fasa kumar terinua*  
ana 1SG-buy pack eight  
'Earlier, I bought eight packs.'
- d. *i-fasa wai\_pipidir lalalak*  
1SG-buy stuff small  
'I buy small stuff.'

The data comes from the first audiovisual corpus of Uruangnirin, which was gathered during a field trip by the author in late 2022. It resulted in 9 hours of transcribed and translated recordings. Of these, 75 minutes are now glossed and further annotated so that differential subject marking can be investigated. The annotated material consists of 14 narratives of six different speakers, and 1427 intonation units. Third-person arguments are excluded from the study because their form makes it difficult to determine whether or not the subject prefix is used after the pronoun, because they have only one prefix form, and to exclude non-human referents. 570 first and second-person argument-verb combinations remain. Table 1 shows the Uruangnirin pronouns and subject indexes.



Table 1. Uruangnirin pronouns and subject indexes

	pronoun	subject index(es)
1SG	<i>lau</i>	<i>lai-, li-, i-</i>
2SG	<i>o</i>	<i>mi-, m-</i>
3SG	<i>ia</i>	<i>a-</i>
1PL.EX	<i>ami</i>	<i>ama-, am-, ma-, mi-, m-</i>
1PL.IN	<i>ita</i>	<i>ta-, t-</i>
2PL	<i>imi</i>	<i>im-, mi-</i>
3PL	<i>sina</i>	<i>a-</i>

Figure 1 shows the variation in subject marking found in the 75-minute corpus.

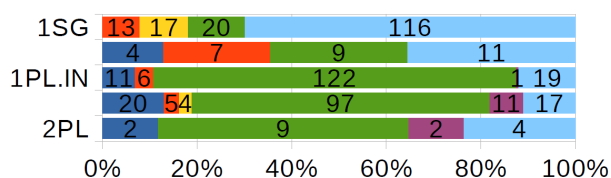


Figure 1: Uruangnirin first and second-person subject marking

The first (quick and dirty) analysis shows that subject marking is not predicted by the transitivity of the verb, the mood of the verb, verb semantics, the semantic role of the subject or by the person or number of the subject (although there are big differences in the distribution of subject markers between the persons and numbers, as is clear from Table 1).

Morphophonology does play a role: the consonant-only subject prefixes (2SG *m-*, 1PL.IN *m-*, 1PL.EX *t-* and 2PL *m-*) only occur on vowel-initial verbs and verbs of the so-called n-class,<sup>1</sup> which are irregularly inflected. However, this only works one way: vowel-initial verbs and verbs of the n-class may also carry longer prefixes or be preceded by a longer prefix and a pronoun.

There are three discourse-related factors that increase the likelihood of the use of a longer form or a pronoun + prefix. These are a long referential distance (Givón 1983), being the first time the argument is mentioned, and a focused argument. Topic persistence (Givón 1983) does not play a role except for in multi-verb constructions: the second verb in such a construction will always carry the shortest possible form. Verbs that are borrowings from Indonesian also tend to carry a longer subject marker. Finally, personal preferences play a role: while one speaker uses the shortest 1SG form *i-* in only 37% of her 1SG argument-verb combinations, another uses no long 1SG prefixes and has *i-* in 100% of the cases. One speaker stands for all the instances of 1PL.EX *am-* and *mi-*, while two others stand for all the instances of *ma-*.

Although these tendencies explain a part of the choices, they are not sufficient to predict all subject marking in Uruangnirin. No variable predicts 100% how the subject is marked. The results presented in this abstract are based on a manual count. At the workshop, I aim to present more systematic counts based on GRAID (Haig and Schnell 2015) and RefIND (Schiborr, Schnell and Thiele 2018) annotations, and to support the claims with a statistical analysis.

## References

Givón, Talmy. 1983. Introduction. In Talmy Givón (ed), *Topic continuity in discourse. A quantitative cross-linguistic study*, 3–41. Amsterdam & Philadelphia: John Benjamins Publishing Company.

- Haig, Geoffrey and Stefan Schnell. 2015. *Annotations using GRAID (Grammatical Relations and Animacy in Discourse)*. Bamberg: Opus.
- Schiborr, Nils N., Stefan Schnell and Hanna Thiele. 2018. *RefIND — Referent Indexing in Naturallanguage Discourse: Annotation guidelines (v1.1)*. Bamberg / Melbourne: University of Bamberg / University of Melbourne.

# A corpus-based study of flexible word order, language contact and language change in Warlpiri

Maria Vollmer

University of Freiburg, Australian National University

This talk presents a corpus-based, usage-oriented diachronic study of flexible word order and its functions in the Australian language Warlpiri (Pama-Nyungan, Central Australia). Warlpiri is spoken in the Northern Territory by about 2,592 people (2021 ABS Census) and is in increasing language contact to English (e.g. Browne 2022; O'Shannessy 2006). The contact to English has led to language change in various areas, including the emergence of the mixed language Light Warlpiri (O'Shannessy 2006). Previous studies have pointed to potential language change in Warlpiri's flexible word order in some communities due to contact to English (Bavin and Shopen 1991; O'Shannessy 2006). Nevertheless, no systematic corpus-based study on potential changes in word order has been conducted so far. Similarly, an extensive, corpus-based study of discourse-functional and structural factors influencing word order variation in Warlpiri spoken discourse is still lacking (e.g. Simpson 2006: 508).

This research project tackles this research gap by creating and analysing the to-date largest morpho-syntactically annotated and reference-tracked corpus of spoken narratives in Warlpiri. The corpus consists of 1,849 clauses, and has been segmented into intonation units (Chafe 1994; Himmelmann et al. 2018), translated, glossed, annotated with GRAID (Haig and Schnell 2014), and reference-tracked (Schiborr et al. 2018). Crucially, the corpus includes some of the earliest available recordings of Warlpiri (from the 1960s; Hale 1966-1967a, b), as well as recordings from the 2000s (Daniels 2009; Morton 2009a-e; Nelson 2009; O'Shannessy 2004a-e, 2022; Presley 2009; Ross 2009), making a comparison between the two time frames with regard to potential language change possible. While the time span of four decades is arguably relatively short for a diachronic study, particularly in comparison with diachronic studies on (written) Indo-European languages, it falls into a time of much change in the Warlpiri communities and increasing contact to and pressure from English. In addition, it opens the possibility of examining language change in spoken discourse in an understudied language.

The talk aims to answer the following research questions: 1) Has word order changed between the 1960s and 2000s, and if yes, has it become more similar to English SVO word order?, and 2) What discourse-functional and structural factors influence variation in word order, and has this changed? Factors include animacy, referential givenness/newness, lexicality, and the frequency of referents measured in different ways. Preliminary results suggest that little change has occurred with regard to subject and verb: SV order is more common than VS order in both time frames, and there has not been much change between older and newer recordings. However, the order of verb and object has changed and VO order has increased in the newer recordings. Since OV order is the more common order in the older recordings, the increase in VO order has paradoxically not led to a more rigid word order like in English, but to an even more flexible word order with a now roughly even split between OV and VO order.

This corpus-based, usage-oriented study takes a first step towards a better understanding of flexible word order and of diachronic processes in an understudied non-European language. The

results therefore have important implications for typological studies as well as language evolution.

## References

- Australian Bureau of Statistics. 2021. Northern Territory. *2021 Census All persons QuickStats*.
- Bavin, Edith and Shopen, Timothy. 1991. Warlpiri in the 80s: An overview of research into language variation and child language. In S. Romaine (ed.), *Language in Australia*, 104-117, Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620881.008
- Browne, Emma. 2022. *Linguistic Innovation and Continuity: Teaching in and of Warlpiri Language at Yuendumu School*. PhD thesis submitted for the degree of Doctor of Philosophy of The Australian National University.
- Chafe, Wallace L. 1994. *Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing*. Chicago and London: The University of Chicago Press.
- Daniels, Dolly Nampijinpa. 2009. Track 6 Yakajirri. Warlpiri-patu-kurlangu Jaru (ed.), *Jaru-kurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.
- Haig, Geoffey and Schnell, Stefan. 2014. Annotations using GRAID (Grammatical Relations and Animacy In Discourse). Manual Version 7.0. In Geoffrey Haig and Stefan Schnell (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. Retrieved 02 August 2019 from <https://multicast.aspra.uni-bamberg.de>.
- Hale, Kenneth L. 1966-1967a. *Warlpiri and Warlmanpa language elicitation*, Australian Institute of Aboriginal and Torres Strait Islander Studies, Call Number HALE\_K01, Tape Number 2.13 Track 2. Recorded at Warrarbi, N.T., and Yuendumu, N.T., 1966-1967, by Ken Hale. Creator: Ken Hale, Performer: Stephen Simpson Tjapangardi, Sam Johnson Tjapangardi, Donald Spencer Tjupurrula. <https://tinyurl.com/2p84u3ez>
- Hale, Kenneth L. 1966-1967b. *Warlpiri language elicitation*, Australian Institute of Aboriginal and Torres Strait Islander Studies, Call Number HALE\_K04, Tape number 2.19. Recorded at Lander River, N.T., and Yuendumu, N.T., 1966-1967, by Ken Hale. Creator: Ken Hale, Performer: Denny Japaljarri, Sam Johnson Japangardi, Paddy Stewart Jupurrula, Weddie Jangala. <https://tinyurl.com/2xxwdpkb>
- Himmelman, Nikolaus, Sandler, Meytal, Strunk, Jan, and Unterladstetter, Volker. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology*, 35(2), 207-245. <https://doi.org/ghmnqf>
- Morton, Lady Napaljarri. 2009a. Track 1 Manja. Warlpiri-patu-kurlangu Jaru (ed.), *Jaru-kurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.
- Morton, Lady Napaljarri. 2009b. Track 2 Janjinngi. Warlpiri-patu-kurlangu Jaru (ed.), *Jaru-kurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.

- Morton, Lady Napaljarri. 2009c. Track 3 Yinapaka. Warlpiri-patu-kurlangu Jaru (ed.), *Jaru- kurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.
- Morton, Lady Napaljarri. 2009d. Track 4 Jitirlpuru. Warlpiri-patu-kurlangu Jaru (ed.), *Jaru- kurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.
- Morton, Lady Napaljarri. 2009e. Track 5 Wapilingki. Warlpiri-patu-kurlangu Jaru (ed.), *Jaru- kurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.
- Nelson, Liddy Nakamarra. 2009. Track 15 Jujuminyi-minyi. Warlpiri-patu-kurlangu Jaru (ed.), *Jarukurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.
- O'Shannessy, Carmel. 2022. Warlpiri DoReCo dataset. In Seifart, Frank, Ludger Paschen and Matthew Stave (eds.). *Language Documentation Reference Corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2). (<https://doi.org/10.34847/nkl.042dv614>).
- O'Shannessy, Carmel. 2004a. Item "ERGstoryWA32\_1" in collection "C. O Shannessy". *The Language Archive*. <https://hdl.handle.net/1839/00-0000-0000-0021-107F-8>. (Accessed 2022-03-21)
- O'Shannessy, Carmel. 2004b. Item "ERGstoryWA32\_2" in collection "C. O Shannessy". *The Language Archive*. <https://hdl.handle.net/1839/00-0000-0000-0021-109B-A>. (Accessed 2022-03-21)
- O'Shannessy, Carmel. 2004c. Item "ERGstoryWA32\_3" in collection "C. O Shannessy". *The Language Archive*. <https://hdl.handle.net/1839/00-0000-0000-0021-10A2-8>. (Accessed 2022-03-21)
- O'Shannessy, Carmel. 2004d. Item "ERGstoryWA29\_1" in collection "C. O Shannessy". *The Language Archive*. <https://hdl.handle.net/1839/00-0000-0000-0021-1092-4>. (Accessed 2022-03-21)
- O'Shannessy, Carmel. 2004e. Item "ERGstoryWA29\_3" in collection "C. O Shannessy". *The Language Archive*. <https://hdl.handle.net/1839/00-0000-0000-0021-1088-6>. (Accessed 2022-03-21)
- O'Shannessy, Carmel. 2006. Language contact and children's bilingual acquisition: Learning a mixed language and Warlpiri in northern Australia. [Doctoral dissertation, University of Sydney]. <http://hdl.handle.net/2123/1303>
- Presley, Molly Napurrurla. 2009. Track 7 Yawakiyi. Warlpiri-patu-kurlangu Jaru (ed.), *Jaru- kurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.
- Ross, Mary Nangala. 2009. Track 17 Ngapa-kurlu. Warlpiri-patu-kurlangu Jaru (ed.), *Jarukurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra),*

*Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD.* Victoria: Trafford Publishing.

Schiborr, Nils, Schnell, Stefan and Thiele, Hanna. 2018. RefIND — Referent indexing in natural-language discourse. Annotation guidelines v1.1. In Geoffrey Haig and Stefan Schnell (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. Retrieved on 02 August 2019 from [multicast.aspra.uni-bamberg.de](http://multicast.aspra.uni-bamberg.de).

Simpson, Jane. 2006. Expressing pragmatic constraints on word order in Warlpiri. In J. Grimshaw, J. Maling, C. Manning, J. Simpson and A. Zaenen (eds.), *Architectures, rules, and preferences: A Festschrift for Joan Bresnan*, 491-514, CSLI Publications. <https://tinyurl.com/3u3bp8f5>