The GRAID-annotated Jinghpaw Corpus: Annotations and Initial Findings

Kurabe, Keita

ILCAA, Tokyo University of Foreign Studies / University of Melbourne

The aim of this paper is two-fold: to explore an implementation of the GRAID glossing conventions to a corpus of Jinghpaw, a Tibeto-Burman language spoken in and around northern Burma; and to give initial findings drawn from the Jinghpaw corpus. Beginning with an overview of clause structures, annotations of referential expressions in terms of forms (NPs, pronouns, zeros), semantics (person and animacy), and functions (S, A, P, obliques, etc.), and types of predicates (verbal, nominal, copula, and non-canonical predicates) are explained. This is followed by a discussion of complex sentences, including complements, relative clauses, adverbial clauses, direct speech, and sentence-level coordination. This paper also explores annotations of constructions with special features, illustrating glossing of external possession, serial verbs, tail-head linkage, and repetition. These annotation conventions are followed by initial findings drawn from our corpus, especially focusing on Preferred Argument Structure, the discourse basis of ergativity, and referential density. Our findings show (a) that the crosslinguistically well-attested low lexicality of A can be replicated by our Jinghpaw data; (b) that the ergative alignment in discourse is not supported by our data; and (c) that approximately half of argumental functions are not overtly expressed in our texts.

1. Introduction

The aim of this paper is two-fold: to outline an implementation of the GRAID glossing conventions (Haig and Schnell 2014) to Jinghpaw (ISO 639-3: kac), a language spoken in northern Burma (Myanmar) and neighboring areas of China and

India; and to give initial findings drawn from the Jinghpaw corpus. The language is affiliated with the Tibeto-Burman branch of the Sino-Tibetan language family. Within Tibeto-Burman, it is closely related to Luish (Asakian) languages such as Cak, Kadu, and Ganan, which are distributed in small discontinuous pockets situated across northwestern Burma, southeastern Bangladesh, and northeastern India. The linguistic data in our Jinghpaw corpus are based on a variety spoken in and around Myitkyina, the Kachin State of Burma, and it is considered to be the standard dialect of the language in Burma. The corpus data consist of traditional narrative texts, all of which are primary data, selected from 1,805 narrative recordings in Jinghpaw. They were collected in Burma by the author and local collaborators between 2009 and 2017 with the help of 196 native narrators. As of January, 2018, 939 stories have been transcribed by the author and native collaborators using the Jinghpaw orthography. All recordings and transcriptions are available online at PARADISEC (the Pacific And Regional Archive for Digital Sources in Endangered Cultures), and named in accordance with the PARADISEC file-naming convention that consists of CollectionID, ItemID, and ContentFile separated by hyphens, for example, KK1-0001-A (Kurabe 2017).

This paper is structured as follows. Beginning with an overview of clause structures in Section 2, Section 3 takes a closer look at Jinghpaw referential expressions with the GRAID glossing in terms of their forms, semantics, and functions. Annotations of predicates, including verbal, nominal, copula, and non-canonical predicates, are given in Section 4. Section 5 explores complex sentences in Jinghpaw, offering an implementation of the GRAID glossing to complement clauses, relative clauses, adverbial clauses, direct speech, and coordinate constructions. Our treatment of constructions with special features, including external possessor constructions, serial verb constructions, tail-head constructions, and repetition, is provided in Section 6. Section 7 gives initial findings drawn from our corpus, especially focusing on Preferred Argument Structure, the discourse basis of ergativity, and referential density.

2. Overview of clause structures

A predicate, both verbal and non-verbal, is always placed at the end of a clause, as illustrated by the following examples with linguistic glosses and GRAID annotations. [1] The two major clause types are verbal clauses headed by a verbal predicate as in (1a), and copula/nominal clauses headed by a nominal predicate with or without a copula as in (1b) and (1c).

(1)   a.   *shi*   *nta*   *=de* *wa*   *=ai*   *=da.*
             3sg   house =ALL return =DECL =HS
        ## pro.h:s np:g =rn v:pred =rv =other
        'She went back home, it's said.' (KK1-0319_051)

---

[1] The list of symbols employed in the GRAID convention is given in the appendix of Schnell and Schiborr (2018).

b.   *ndai*   *ngai*   *=na*   *nga*   *nan*   *re.*
         this   1sg   =GEN   COW   exactly   COP
   ##ds   dem_pro:s   pro.1:poss   =rn   np:pred   other   cop

'This is exactly my cow.' (KK1-0272_040)

c.   *ndai*   *ngai*   *=na*   *nga.*
         this   1sg   =GEN   COW
   ##ds   dem_pro:s   pro.1:poss   =rn   np:pred

'This is my cow.' (elicited)

All nominal expressions, excluding afterthoughts, occur before predicates, as shown above. NPs, especially non-core arguments, are marked by postpositive case marking clitics to indicate their relationship to the predicate. The order of NPs, as seen in (2a) and (2b), is relatively free, being determined by pragmatic factors. Because predicates are the only obligatory constituents of clauses, references of arguments, when pragmatically retrievable from the context, are freely omitted in Jinghpaw, as seen in (2c).

(2)   a.   *nang*   *ngai*   *=hpe*   *grau*   *tsawra*   *=ai*   *=i?*
             2sg   1sg   =ACC   more   love   =DECL   =Q
     ##ds   pro.2:a   pro.1:p   =rn   other   v:pred   =rv   =other

'Do you love me more (than him)?' (KK1-0262_066)

b.   *nanhte*   *=hpe*   *ngai*   *tsawra*   *=nga*   *=na*   *=yaw.*
         2pl   =ACC   1sg   love   =CONT   =IRR   =SFP
   ##ds   pro.2:p   =rn   pro.1:a   v:pred   =aux   =aux   =other

'I will love you (always).' (KK1-0474_026)

c.   *nta*   *duhkra*   *hkan*   *shachyut*   *=ai*   *=da.*
                   house   TERM   follow   chase   =DECL   =HS
   ##   0.d:a   0.h:p   np:g   other   lv   v:pred   =rv   =other

'(The spirit) chased (her) to her house, it's said.' (KK1-0319_066)

3. Referential expressions

Glossing of referential expressions is a fundamental part of GRAID annotations. This section, following Haig and Schnell (2014), explores Jinghpaw referential expressions in terms of their forms (3.1), semantics (3.2), and functions (3.3).

3.1. Form of referential expressions
The distinction between NP, pronoun, and zero is captured by the form glosses ⟨np⟩, ⟨pro⟩, and ⟨0⟩, respectively.

3.1.1. NP

In GRAID annotations, the form gloss ⟨np⟩ is given to what in the literature is labeled "lexical mention/expression," and so on (Du Bois 1987, Haig and Schnell 2014). As with other corpora, the most typical NPs glossed with the form gloss ⟨np⟩ in our corpus are those headed by common nouns. NPs are also headed by kinship terms, person and place names, and so on. These NPs are introduced with or without case-marking postclitics depending on their functions and/or semantic roles (see 3.3). Examples:

(3)　a.　　*ndai*　　*u-hka*　　*=mung*　*mau*　　　　*=mat*　　*=na*　　*=she...*
　　　　　　this　　　bird-crow　=also　　be.surprised　=COMPL　=SEQ　=then
　　　##　ln_dem　np.d:s　=other　v:pred　　　=aux　　=other　=other
　　　　'This crow was also surprised and...' (KK1-0275_032)

　　b.　　　　　　　*ganu*　　*=hpe*　*bai*　　*yu*　　*=dat*　　*=ai*　*shaloi*
　　　　　　　　　　mother　=ACC　again　see　　=away　=NMLZ　when
　　　#ac　0.h:a　np.h:p　=rn　　other　v:pred　=aux　　=rv　　other
　　　=wa...
　　　=TOP
　　　=other
　　　'When (he) saw his mother again...' (KK1-0187_024)

Numerals also receive the form gloss ⟨np⟩ when they head NPs. Unlike neighboring classifier languages, such as Shan and Burmese, Jinghpaw numerals, displaying a similar distribution of common nouns, can occur in the absence of semantic heads and classifiers. Numerals, as such, can express different types of referents. The numeral *masum* 'three,' for instance, may denote 'three persons,' 'three dogs,' 'three houses,' 'three books,' and so on, unlike other languages in the region.

(4)　　*wora*　　*masum*　*=gaw*　*yup*　　*rawt*　　*=ai*　　*=hte...*
　　　that　　　three　　=TOP　　sleep　awake　=NMLZ　=COM
　　##　ln_dem　np.h:s　=other　lv　　　v:pred　=rv　　=rn
　　　'The three (men) wake up and...' (KK1-0271_038)

Interrogative pro-forms are also glossed with the form gloss ⟨np⟩ when they head NPs. Interrogatives, often followed by additive particles meaning 'also,' can be employed to express indefiniteness as well. The relationship between major interrogative and indefinite meanings is summarized in Table 1:

Table 1 The interrogative-indefinite relationship

categories	forms	interrogative	indefinite	negative indefinite
person	gadai	who	anybody	nobody
thing	hpa	what	anything	nothing
place	gara	where	anywhere	nowhere
amount	gade	how many	any amount	no amount
time	galoi	when	anytime	never
manner	ganing	how	anyhow	no way

Examples follow:

(5) a.

		ma	*=e*	*gadai*	*=ni*	*=hte*	*sa*
		child	=SFP	who	=PL	=COM	go
##ds	0.h:s	np.h:voc	=other	np.h:obl	=rn	=rn	v:pred

=na	*=ma.*
=IRR	=Q
=aux	=other

'Child, who (pl.) will (you) go with?' (KK1-0269_028)

b.

	shi	*=mung*	*hpa*	*=ma*	*n-chye*	*=ai*	*=le.*
	3sg	=also	what	=also	NEG-know	=DECL	=SFP
##neg	pro.h:a	=other	np:p	=other	v:pred	=rv	=other

'He also knows nothing.' (KK1-0269_213)

For complex NPs, the form gloss ⟨np⟩ is given to their heads, and other NP-internal subconstituents, excluding possessors that are specified for their own functions (see 3.4), are glossed as ⟨ln⟩ or ⟨rn⟩ depending on their relative positions to the head. Typical NP-internal subconstituents in our corpus include: adnominal demonstratives, numerals, adjectives, and the pluralizing postclitic *ni*, which encodes additive, collective, and associative plural meanings. Examples:

(6) a.

	dai	*la*	*masum*	*dai*		*=ni*	*=gaw...*
	that	man	three	that		=PL	=TOP
##	ln_dem	np.h:dt_s	rn	dem_pro.h:s		=rn	=other

'These three men, they (are laborers and)...' (KK1-0271_011)

b.

		nga	*gaba*	*=ni*	*n-lu*	*rim*	*=ai*
		fish	big	=PL	NEG-be.able	catch	=NMLZ
#ac.neg	0.h:a	np:p	rn	=rn	lv	v:pred	=rv

majaw...
because...
other

'Because (she) could not get large fish...' (KK1-0187_021)

In conjunctive coordination of NPs, whether monosyndetic (i.e., [A-co] [B]) or

disyndetic (i.e., [A-co] [B-co]), the form gloss ⟨np⟩ is given to the last coordinant in our corpus, and other coordinants to its left are glossed with the form gloss ⟨ln⟩. Jinghpaw has various kinds of coordinators, including comitative case postclitic *hte* 'with, and,' additive particle *mada* 'also,' and a special coordinator *yen* that is restricted to binary coordination of human beings.

(7)   a.

	makaw	*yen*	*magam*	*daini*	=*gaw*	*nan*
	firstborn.daughter	and	firstborn.son	today	=TOP	2du
##	ln	rn	np.d:voc	other	=other	pro.2:s

nau...
sibling
np.d:appos

'Firstborn daughter and firstborn son, today, you siblings...' (KK1-0263_017)

b.

	shannga	=*ni*	=*gaw*	*ayi*	*mada*	*ala*	*mada*	*nrung*
	deer	=PL	=TOP	female	also	male	also	horn
##	np.d:dt	=rn	=other	ln	rn	np.d:appos	rn	np:s

	grai	*tsawm*	=*na...*
	very	be.beautiful	=SEQ
	other	v:pred	=other

'As for deer, both male and female, their horns were very beautiful and...' (KK1-0263_002)

3.1.2. Pronoun

Our corpus, following Haig and Schnell (2014: 9) who intend to capture "definite pronouns," labels personal and demonstrative pronouns with the form gloss ⟨pro⟩. Examples of personal pronouns include:

(8)   a.

	ngai	*nang*	=*hpe*	*garum*	*hpyi*	=*mayu*	=*ai.*
	1sg	2sg	=ACC	help	ask	=DESID	=DECL
##ds	pro.1:a	pro.2:p	=rn	lv	v:pred	=aux	=rv

'I want to ask you for help.' (KK1-0276_067)

b.

	shi	=*gaw*	*dai*	=*ni*	=*hpe*	*hta*	*la*
	3sg	=TOP	that	=PL	=ACC	pick	take
##	pro.h:a	=other	dem_pro:p	=rn	=rn	lv	v:pred

=*na*	=*she...*
=SEQ	=then
=other	=other

'He picked and took these (fish) and then...' (KK1-0269_067)

Jinghpaw personal pronouns are encoded as free pronouns with full forms whose systems exhibit three-way splits in person (1st, 2nd, 3rd) and in number (singular, dual, plural), yielding the paradigm given in Table 2. All personal pronouns, as can be seen, are formally distinguished, and there is no syncretism of person or number

contrasts. The dual/plural distinction is only found in personal pronouns. Separate genitive forms exist for singular, which have their diachronic sources in a contraction of singular personal pronouns plus the obsolete genitive case marker *a*.

Table 2   Personal pronouns

	sg (NOM)	sg (GEN)	dual	plural
1st	ngai	nye	an	anhte
2nd	nang	na	nan	nanhte
3rd	shi	shi	shan	shanhte

Demonstratives, when they head NPs, also receive the form gloss ⟨pro⟩. They are glossed with an additional gloss ⟨dem⟩, which is combined with other glosses separated by an underscore, for example, ⟨dem_pro.h:s⟩. Demonstratives in Jinghpaw function as free pronouns, and distinguish speaker-addressee orientation, relative distance, or relative height from the deictic center: *ndai* [proximate; speaker-centered]; *dai* [proximate; addressee-centered]; *htora* [distal; up]; *wora* [distal; level]; *lera* [distal; down]. Demonstratives, in terms of qualitative features of the referent, are specified for inanimate by default, as in (9a), unless followed by the plural marker *ni*, which turns demonstratives into neutral for humanness and animacy, as in (9b).

(9)   a.   *shi*   *=gaw*   *dai*   *=hpe*   *lang*   *di*   *=na...*
            3sg   =TOP   that   =ACC   hold   LV   =SEQ
       ##   pro.d:a   =rn   dem_pro:p   =rn   v:pred   other   =other
       'He held it (meat) and...' (KK1-0275_010)

      b.   *ndai*   *=ni*   *=gaw*   *ngai*   *=hpe*   *gara*   *=hku*
            this   =PL   =TOP   1sg   =ACC   how   =like
       #ds dem_pro.h:a =rn   =other   pro.1:p   =rn   other   =rn
            wa   *masusha*   *=na*   *=i?*
            return   deceive   =IRR   =Q
            lv   v:pred   =aux   =other
       'How will these ones come back and lie to me?' (KK1-0271_050)

Complex NPs headed by pronouns, including personal and demonstrative pronouns, show reduced possibilities for their internal subconstituents. Unlike common nouns and numerals, they cannot take multiple types of adnominals, such as adjectives, demonstratives, genitives, and relative clauses. Typical complex NPs with pronoun heads in our corpus, as in (9b), are those with numerals and pluralizing markers.

3.1.3. Zero
Almost all references of arguments, when pragmatically recoverable from the context, can be freely omitted in Jinghpaw. In our corpus, zero arguments are assumed

when they are: (a) licensed by the argument structure of a verb; (b) recoverable from the discourse; and (c) not constructionally suppressed. Usually, these zero arguments can alternatively be expressed by overt forms. Zero arguments receive the form gloss ⟨0⟩. Because all NPs, as noted in Section 2, occur before predicates and their order is relatively free, it is impossible to determine the exact position of zero arguments. Instead of arbitrarily determining their positions, we put all of them at the beginning of clauses regardless of their syntactic functions.

(10)   a.   *shi*   *=hpe*   *la*   *=taw*   *=ai.*
                     3sg   =ACC   wait   =CONT   =DECL
       ## 0.h:a   pro.h:p   =rn   v:pred   =aux   =rv
       '(She) was waiting for him.' (KK1-0187_060)

       b.   *nang*   *=gaw*   *sa*   *rim*   *=u.*
                     2sg   =TOP   go   catch   =IMP
       ##ds 0:p   pro.2:a   =other   lv   v:pred   =rv
       'You go catch (fish)!' (KK1-0276_017)

       c.   *hpai*   *=wa*   *=na*   *=she...*
                     carry   =VEN   =SEQ   =then
       ## 0.h:a 0:p   v:pred   =aux   =other   =other
       '(He) carried (the fish) back home and then...' (KK1-0269_071)

We assume a zero in the S position of imperative and hortative clauses, despite them often being omitted, as in (11a) and (11b), because they can alternatively be expressed by overt forms, as in (11c) and (11d). For cases where we do not assume a zero, see Section 4.3.

(11)   a.   *atsawm* *rai*   *=na*   *bai*   *wa*   *=u*   *=yaw.*
                     well   LV   =SEQ   again   return   =IMP   =SFP
       ##ds 0.2:s   other   other   =other   other   v:pred   =rv   =other
       '(You) come back carefully!' (KK1-0269_038)

       b.   *mare*   *gaba*   *=de*   *sa*   *=mat*   *=ga.*
                     village big   =ALL   go   =COMPL   =HORT
       ##ds 0.1:s   np:g   rn   =rn   v:pred   =aux   =rv
       'Let (us) go to a large village.' (KK1-0262_058)

       c.   *nang*   *atsawm* *rai*   *=na*   *bai*   *wa*   *=u*
                     2sg   well   LV   =SEQ   again   return   =IMP
       ##ds   pro.2:s   other   other   =other   other   v:pred   =rv
       =yaw.
       =SFP
       =other
       'You come back carefully!' (elicited)

d.

anhte	*mare*	*gaba*	*=de*	*sa*	*=mat*	*=ga.*
1pl	village	big	=ALL	go	=COMPL	=HORT
##ds pro.1:s	np:g	rn	=rn	v:pred	=aux	=rv

'Let us go to a large village.' (elicited)

3.2. Animacy and person of referential expressions

Four animacy and person glosses, that is, ⟨h⟩, ⟨d⟩, ⟨1⟩, and ⟨2⟩, are considered in the Jinghpaw corpus. Reference to speech-act participants, glossed with ⟨1⟩ and ⟨2⟩, only occur in direct speech in our corpus because it consists of narrative texts. Examples include:

(12) a.

shi	*ganu*	*=hpe*	*grai*	*tawngban*	*=ai*	*=i.*
3sg	mother	=ACC	very	apologize	=DECL	=SFP
## pro.h:a	np.h:p	=rn	other	v:pred	=rv	=other

'He apologized to his mother a lot, OK?' (KK1-0187_066)

b.

nang	*ngai*	*=hpe*	*grau*	*tsawra*	*=ai*	*=i?*
2sg	1sg	=ACC	more	love	=DECL	=Q
##ds pro.2:a	pro.1:p	=rn	other	v:pred	=rv	=other

'Do you love me more (than him)?' (KK1-0262_066)

Because our corpus contains some fables that feature animals and spirits, the animacy gloss ⟨d⟩ is also employed to gloss anthropomorphized discourse participants. These referents are given human qualities, such as the ability to speak human language. Examples:

(13) a.

shi	*=gaw*	*shi*	*=na*	*gasha*	*=hpe*	*tsun*
3sg	=TOP	3sg	=GEN	child	=ACC	say
## pro.d:a	=other	pro.d:poss	=rn	np.d:p	=rn	v:pred

=ai.
=DECL
=rv

'He (deer) said to his children.' (KK1-0263_016)

b.

aba	*ngai*	*wora*	*maling*	*=de*	*agu*	*wadu-brang*
brother	1sg	that	forest	=ALL	uncle	boar-brother
##ds np.d:voc	pro.1:s	ln_dem	np:g	=rn	ln	ln

=ni	*dumsi*	*=ni*	*=kaw*	*sa*	*lam*	*=na.*
=PL	porcupine	=PL	=LOC	go	wander	=IRR
=rn	np.d:g	=rn	=rn	lv	v:pred	=aux

'Brother (deer), I will go wander to the forest, to uncle boars, porcupines, and others.' (KK1-0263_031)

Animacy features are given depending on semantic role consideration of the referent rather than its form. Thus, the same noun may be assigned different animacy values

depending on its meaning in context. For example, a place noun *mare* 'village' is treated as a goal and receives no animacy feature in (14a), in contrast to (14b), where the same noun is metonymically used for those who live in the village.

(14) a.   *mare   langai mi   =kaw du   yang   =gaw...*
              village   one   one   =LOC arrive   when   =TOP
         ## 0.h:s np:g   rn   rn   =rn   v:pred   other   =other
         '(They) arrived at one village and...' (KK1-0274_011)

     b.   *mare   ting   nga-nawng jawm   htawk   =ai   %*
              village   whole   fish-pond do.together   remove   =NMLZ
         #rc np.h:a   rn   np:p   lv   v:pred   =rn   %
         nhtoi-lamang...
         day-program
         np:p
         'a (festival) day program where the villagers removed water from a pond (to catch fish)...' (KK1-0187_008)

3.3. Function of referential expressions
3.3.1. Core argument function

NPs, as noted in Section 2, may be marked by postpositive case marking clitics. The case marking pattern, as shown in (15a) and (15b), is the nominative-accusative type (S/A vs. P), where the S and A functions occur without any overt case marker in contrast to the P, which may be case-marked by an accusative postclitic. The P function, as seen in (15c), may also be introduced without any overt marker (see below).

(15) a.   *ngai   sa   =na   =law.*
              1sg   go   =IRR   =SFP
         ##ds pro.1:s v:pred   =aux   =other
         'I will go.' (KK1-0269_035)

     b.   *ngai   nang   =hpe garum hpyi   =mayu =ai.*
              1sg   2sg   =ACC help   ask   =DESID =DECL
         ##ds pro.1:a pro.2:p   =rn   lv   v:pred   =aux   =rv
         'I want to ask you for help.' (KK1-0276_067)

     c.   *raitimung nang   hpaji   naw   ra   =ai   =yaw.*
              but   2sg   knowledge   still   need   =DECL   =SFP
         ##ds other   pro.2:a   np:p   other   v:pred   =rv   =other
         'But, you still need knowledge.' (KK1-0275_036)

The P function is obligatorily marked by the accusative when the P outranks or is equal to the A on the animacy hierarchy given in (16), and is left unmarked or optionally marked by the accusative when the P is lower than the A. This case marking pattern is well motivated by the need to differentiate between two potential agents by overtly

case marking the P with the accusative, leaving the A unmarked (Kurabe 2012). In other words, the P is case marked when there is a possibility that it may be misconstrued with the A otherwise, the situation of which arises when the P is equally high or higher than the A on the animacy hierarchy, as the prototypical P is lower than the A in animacy (Comrie 1981: 121).[2]

(16) Animacy hierarchy (Comrie 1981: 178)
      human > animal > inanimate

A similar case marking pattern is also observed for ditransitive constructions whose case frame is that the A remains unmarked, the recipient is obligatorily marked by the accusative, and the theme is left unmarked. This is due to the fact that the recipient is typically human; this is in contrast to the theme, which is typically non-human, and ranked lower than the agent and recipient on the animacy hierarchy.[3] To illustrate this, consider (17). In our corpus, the theme is glossed as ⟨:p2⟩.

(17)   *nang*   *ngai*   =*hpe* *manu-jahpu* *jaw*   =*na...*
              2sg   1sg   =ACC   price-price   give   =IRR
      ##ds   pro.2:a   pro.1:p   =rn   np:p2   v:pred   =aux
      'You would give me the price...' (KK1-0276_088)

3.3.2. Oblique function

Obliques are encoded by means of case-marking postclitics. As core arguments, obliques always precede verbal predicates with relatively free orders. Following the GRAID manual (Haig and Schnell 2014: 14–6), our corpus considers three types of adjuncts: locations ⟨:l⟩, goals ⟨:g⟩, and other semantic roles ⟨:obl⟩. Locations and goals can both be encoded by the locative case *kaw* that marks physical locations (18a), goals (18b), and sources (18c) (in the last case, with an ablative case). Examples follow:

(18)   a.   *ndai*   =*kaw* *jahkring*   *mi*   *naw*   *hkring*
                     this   =LOC   for.a.while   one   still   rest
         ##ds 0.1:s dem_np:l   =rn   other   rn   other   lv
         la   =*ga.*
         take   =HORT
         v:pred   =rv
         'Let's take a rest here for a while.' (KK1-0271_019)

---

[2] This type of case marking employed for disambiguation of the P from the A is widespread among TB languages as well as is crosslinguistically common (LaPolla 1992, Malchukov 2008, and others). The definitness of the P function, although known to play some role in some languages, does not play a role in Jinghpaw. This can be seen in the fact that P arguments low in definiteness, such as interrogatives, can potentially be marked with the accusative.
[3] When equally-ranking recipient and theme NPs are involved, both of them must be case marked with the accusative. In such situations, the interpretation of the recipient and theme is determined by context, as scrambling of the recipient and theme NPs does not contribute to the meaning (Kurabe 2012).

b. *shi*   *=gaw*   *lup*   *=kaw*   *du*   *=ai*   *=hte*   *=i...*
    3sg   =TOP   grave   =LOC   arrive   =NMLZ   =COM   =SFP
pro.h:s   =other   np:g   =rn   v:pred   =rv   =rn   =other
'He arrived at the grave and, OK?' (KK1-0274_045)

c.   *ma*   *nang*   *ndai*   *gara*   *=kaw*   *=na*   *la*
       child   2sg   this   where   =LOC   =ABL   take
##ds   np.h:voc   pro.2:a   dem_np:p   np:l   =rn   =rn   v:pred
=wa   *=ai*   *=rai?*
=VEN   =DECL   =Q
=aux   =rv   =other
'Child, where did you take this from?' (KK1-0269_204)

Goals can also be marked by the allative case *de*, as in (19a), which, unlike the locative *kaw*, is sensitive to the animacy of the goal NP it marks: it can only mark inanimate goals, for example, *nta de* 'to the house' vs. **manang de* 'to the friend.' The allative can mark animate goals only by means of *hpang* (originally a locator noun meaning 'behind'), which is employed to "locationalize" animate nouns, as in (19b).

(19) a.   *mare*   *gaba*   *=de*   *sa*   *=mat*   *=ga.*
                              village   big   =ALL   go   =COMPL   =HORT
      ##ds 0.1:s   np:g   rn   =rn   v:pred   =aux   =rv
'Let's go to a large village.' (KK1-0262_058)

b.   *shanhte*   *=ni*   *hkawhkam-wa*   *hpang*   *=de*   *wa*
                        3pl   =PL   king-man   LOCATIONALIZER   =ALL   go
0.h:p   pro.h:a   =rn   np.h:g   rn   =rn   lv
sa   *=ya*   *=na*   *=she...*
send   =BEN   =SEQ   =then
v:pred   =aux   =other   =other
'They sent him to the king and then...' (KK1-0269_197)

Other obliques receive the function gloss ⟨:obl⟩, marking semantic roles, such as companion (20a), instrument (20b), and so on.

(20) a.   *gwi*   *=hte*   *rau*   *hpun*   *hta*   *sa*   *=na*
                              dog   =COM   together   wood   pick   go   =SEQ
      ## 0.h:a   np:obl   =rn   other   np:p   lv   v:pred   =other
=she...
=then
=other
'(He) went to pick up wood with his dog and then...' (KK1-0269_113)

b.   *shupsheng*   *=hte*   *bau*   *=ni*   *=hte*   *shangoi*
                        cymbal   =COM   drum   =PL   =COM   make.a.noise
0.h:s   ln   =rn   np:obl   =rn   =rn   v:pred

=*na...*
=SEQ
=other
'(They) made a noise with cymbals and drums and...' (KK1-0269_234)

Circumstantial adjuncts, even when marked by the locative case *kaw* like locations, are given the function gloss ⟨:other⟩ in accordance with the GRAID manual (Haig and Schnell 2014: 17), where obliques are glossed depending on semantic role considerations rather than their forms.

3.4. Other syntactic functions

Other glosses for syntactic functions considered in the Jinghpaw corpus include: ⟨:dt⟩ for dislocated topics, ⟨:voc⟩ for vocatives, ⟨:appos⟩ for appositionals, and ⟨:poss⟩ for possessors. The gloss ⟨:dt⟩ is given to NPs that occur outside clause boundaries. No distinction is made between right and left dislocation. The function of clause-internal coreferential elements, when relevant, is also marked for dislocated phrases.

(21) a.   *dai*   *mare*   *langai*   *mi*   *ndai*   =*kaw...*
             well   village   one   one   this   =LOC
       ##   other   np:dt_l   rn   rn   dem_np:l   =rn
       'Well, at one village, there...' (KK1-0262_002)

      b.   *nang*   =*da*   #ac   *nang*   *gahpu*   *re*
                 2sg   =HS   2sg   elder.brother   COP
       ##ds   pro.2:dt_s   =other   #ac   pro.2:s   np.h:pred   cop
       majaw...
       because
       other
       'You, she said, because you are the elder brother...' (KK1-0262_013)

The function gloss ⟨:voc⟩ is applied to vocative phrases, which are typically kinship terms. Examples:

(22) a.   *adwi*   *n-matsing*   =*s-ai.*
                                          grandmother   NEG-remember   =CSM-DECL
       ##ds.neg 0.1:a 0:p   np.d:voc   v:pred   =rv
       'Grandma, (I) don't remember (it) anymore.' (KK1-0269_091)

      b.   *ma*   *nang*   *ndai*   *gara*   =*kaw* =*na*   *la*
             child   2sg   this   where   =LOC   =ABL   take
       ##ds np.h:voc   pro.2:a   dem_np:p   np:l   =rn   =rn   v:pred
       =*wa*   =*ai*   =*rai?*
       =VEN   =DECL   =Q
       =aux   =rv   =other
       'Child, where did you take this back from?' (KK1-0269_204)

The function gloss ⟨:appos⟩ is given to appositional phrases, which are co-referent with juxtaposed phrases, adding additional information to the referent.

(23) a.   *anhte*   *shinggyim-masha* *=ni*   *=gaw...*
                1pl   human-person   =PL   =TOP
       ##ds   pro.1:s   np.h:appos   =rn   =other
       'We, human beings, (are)...' (KK1-0319_004)

      b.   *shi*   *=na* *gasha*   *magam-wa*   *=hpe*
                            3sg   =GEN   child   firstborn-man =ACC
       ## 0.h:a 0:p2   pro.h:poss   =rn   np.h:p   np.h:appos   =rn
       jaw   *=ai.*
       give   =DECL
       v:pred   =rv

       '(The mother) gave (it) to her son, to the firstborn son.' (KK1-0262_012)

Possessors, which are typically encoded with the genitive case *na*, are glossed with the functional gloss ⟨:poss⟩, which is the only NP-internal function in GRAID. Separate genitive forms, as noted in 3.1.2, exist for singular personal pronouns, for example, *na* 'your (sg.).'

(24) a.   *ndai*   *gumra* *=wa*   *anhte*   *=na*   *mam-nli* *mahkra*
                this   horse   =TOP   1pl   =GEN   rice-seed all
       ##ds   ln_dem   np:a   =other   pro.1:poss   =rn   np:p   rn
       sha   *=kau*   *=ya*   *=s-ai.*
       eat   =away   =BEN   =CSM-DECL
       v:pred   =aux   =aux   =rv
       'This horse ate all of our rice seeds.' (KK1-0271_057)

      b.   *na*   *kashu*   *she*   *rai*   *=s-ai*
                            2sg.GEN   grandson   indeed   COP   =CSM-DECL
       ##ds 0.h:s   pro.2:poss   np.h:pred   other   cop   =rv
       =gaw.
       =SFP
       =other
       '(He) is indeed your grandson.' (KK1-0269_250)

4. Predicates

This section provides the GRAID glossing of predicates in the Jinghpaw corpus, beginning with verbal predicates (4.1), followed by copula/nominal (4.2), and non-canonical predicates (4.3).

4.1. Verbal predicates

Predicates headed by verbs receive the form gloss ⟨v⟩ and function gloss ⟨:pred⟩. The copula, although it is morphosyntactically a verb in Jinghpaw, is glossed differently with the gloss ⟨cop⟩ (see 4.2). Jinghpaw is an aspect- and mood-prominent language with no grammatical tense. Verbs are typically followed by mood-marking postclitics consisting of six paradigmatic values, for example, *ai* 'DECL' and *u* 'IMP,' which mark the end of the verbal predicate. Mood markers, as illustrated by (25), are glossed with the form gloss ⟨rv⟩. Elements occurring after mood markers, such as sentence-final particles, are elements outside the verbal predicate, and thus receive the form gloss ⟨other⟩, as in (25b).

(25) a.
shi	*gasha*	*=hpe*	*grai*	*tsawra*	*=ai.*
3sg	son	=ACC	very	love	=DECL
## pro.h:a	np.h:p	=rn	other	v:pred	=rv

'She loves her son very much.' (KK1-0187_033)

b.
raitimung	*nang*	*hpaji*	*naw*	*ra*	*=ai*	*=yaw.*
but	2sg	knowledge	still	need	=DECL	=SFP
##ds other	pro.2:a	np:p	other	v:pred	=rv	=other

'But, you still need knowledge.' (KK1-0275_036)

Jinghpaw verbs fall into two primary lexical aspect classes: the active verb, which semantically encodes a dynamic situation or 'something happens,' and the stative verb, which encodes a stative situation or a non-happening. The importance of this classification primarily lies in the temporal interpretation of verbs with the declarative mood marker *ai*. When followed by this marker, an active verb, as in (26a), normally indicates the time prior to the moment of speech, while a stative verb, as in (26b), normally indicates the present moment (although time reference is changeable with the help of temporal adverbs, such as *shani shagu* 'every day' and *moi* 'long ago'). Both active and stative verbs receive the gloss ⟨v:pred⟩ in our corpus.

(26) a.
shi	*=gaw*	*dai*	*=kaw*	*yup*	*=ai.*
3sg	=TOP	that	=LOC	sleep	=DECL
## pro.h:s	=other	dem_np:l	=rn	v:pred	=rv

'He slept there.' (KK1-0274_052)

b.
jan	*grai*	*ja*	*=ai*	*=law.*
sun	very	be.hot	=DECL	=SFP
##ds np:s	other	v:pred	=rv	=other

'It's very hot.' (KK1-0271_018)

Many morphosyntactic properties show that words denoting property concepts, such as *ja* 'be hot,' can be best treated as stative verbs in Jinghpaw, being thus glossed with ⟨v:pred⟩. Note further that some stative verbs, especially those denoting the four

core semantic types of adjectives (Dixon 1977), unlike other stative verbs, have an additional ability to modify nouns in the post-nominal position without any marker of syntactic dependency. We label them as "adjectives" and treat them as a subclass of stative verbs. When functioning as predicates, adjectives receive the gloss ⟨v:pred⟩, and when functioning as modifiers, they receive the gloss ⟨rn⟩. Compare:

(27) a. *ndai*   *nga*   *=ni*   *gaba*   *=ai.*
          this   fish   =PL   be.big   =DECL
      ## ln_dem np:s =rn   v:pred =rv
          'These fish are big.' (elicited)

      b.   *nga*   *gaba* *=ni* *n-lu*   *rim*   *=ai*   *majaw...*
                      fish   big   =PL   NEG-get   catch   =NMLZ   because
      #ac.neg 0.h:a np:p rn   =rn   lv   v:pred =rv   other
          'Because (she) could not get big fish...' (KK1-0187_021)

Verbs may be followed by an array of optional auxiliaries, expressing meanings associated with aspectuality, modality, evidentiality, intensity, and so on. Auxiliaries are glossed with the form gloss ⟨aux⟩ in our corpus.

(28) a.   *shi*   *=hpe* *la*   *=taw*   *=ai.*
                      3sg   =ACC   wait   =CONT   =DECL
      ## 0.h:a pro.h:p =rn   v:pred =aux =rv
          '(She) was waiting for him.' (KK1-0187_060)

      b. *dai*   *ma*   *=gaw*   *manang*   *=ni* *=hte* *grai*
          that   child   =TOP   friend   =PL   =COM   very
      ## ln_dem np.h:s =other np.h:obl =rn   =rn   other
          chyai   *=mayu*   *=ai.*
          play   =DESID   =DECL
          v:pred =aux   =rv
          'The child really wants to play with friends.' (KK1-0269_016)

4.2. Verbless and copula predicates

Verbless predicates typically consist of nominal predicates. The relation, encoded by a nominal predicate, may be identity (equation), classification, and location, as illustrated by the following examples. The nominal predicate receives the gloss ⟨np:pred⟩.

(29) a. *shi*   *=gaw*   *anhte*   *=na* *manang.*
          3sg   =TOP   1pl   =GEN   friend
      ## pro.h:s =other pro.1:poss =rn   np:pred
          'He is our friend.' (elicited)

b.   *dai*   *namlap*   *dai*   =*gaw*   *ndai*
        that   leaf   that   =TOP   well
    ##   ln_dem   np:dt_s   dem_pro:s   =other   other
       tsihkrungtsinan   =*i.*
       elixir.of.immortality   =SFP
       np:pred   =other
       'The leaf is an immortality elixir, OK?' (KK1-0269_086)

c.   *shi*   =*gaw*   *ya*   *nta*   =*kaw.*
        3sg   =TOP   now   house   =LOC
    ##   pro.h:s   =other   other   np:l:pred   =rn
       'He is at home now.' (elicited)

Nominal predicates show reduced morphosyntactic possibilities. They cannot express properties associated with verbs. For example, they cannot be negated, cannot be specified for aspect and mood, cannot be elaborated by auxiliaries, and cannot be modified by adverbs. These morphosyntactic properties must be encoded by means of a verb, in this case, the copula verb. All relations encoded by a nominal predicate, as shown below, can also be expressed with a copula. A copula, which has the function to relate the subject of a clause with a copula complement, receives a special gloss ⟨cop⟩. In a copula clause, the copula complement always follows the copula subject. This is in contrast to a transitive clause, which also takes two core arguments, but they have a flexible order.

(30)   a.   *ndai*   *ngai*   =*na*   *nga*   *she*   *re*
           this   1sg   =GEN   cow   indeed   COP
     ##ds   dem_pro:s   pro.1:poss   =rn   np:pred   other   cop
           =*gaw.*
           =SFP
           =other
           'This is indeed my cow.' (KK1-0272_033)

     b.   *nat-ga*   *re*   %   *nga*   =*na...*
               spirit-language   COP   say   =SEQ
     ##ds ∅:s   np:pred   cop   %   v:pred   =other
           '(They) said "(it) is a spirit language" and...' (KK1-0319_017)

     c.   *shi*   =*gaw*   *ya*   *nta*   =*kaw*   *re.*
           3sg   =TOP   now   house   =LOC   COP
     ##   pro.h:s   =other   other   np:l:pred   =rn   cop
           'He is at home now.' (elicited)

4.3. Non-canonical predicates

Predicates that exhibit reduced possibilities for government of verbal arguments receive the gloss ⟨vother:pred⟩ (Haig and Schnell 2014: 22–3). Three predicates

heading dependent clauses fall into this category in the Jinghpaw corpus, all involving the core functions S and A. Because arguments are systematically suppressed, no zeros are assumed in the glossing for these cases. The first example comes from a predicate with the subordinator *let*, which forms a simultaneous adverbial clause. One constraint imposed on this construction is that the S or A argument in the dependent clause, which is always coreferential with the S or A argument in the main clause, must not be overtly expressed. Consider:

(31)	*ganu*	*=gaw*	#ac	*grai*	*matsan*	*=let*	%	*=sha*
		mother	=TOP		very	be.poor	=SIM		=only
##	np.h:a	=other	#ac	other	vother:pred	=other	%	=other

grai	*gasha*	*=hpe*	*baumaka*	*=ai.*
very	child	=ACC	take.care.of	=DECL
other	np.h:p	=rn	v:pred	=rv

'The mother, while being very poor, took care of her son very much.'
(KK1-0187_005)

The next example comes from a negated nominalized clause that forms an adverbial clause conveying the sense of privation. Again, the S or A in the dependent clause, being coreferential with the S or A in the main clause, is systematically suppressed.

(32)	*shi*	#ac	*nba*	*n-hpun*	*=ai*	%	*=sha*
		3sg		blanket	NEG-wear	=NMLZ		=only
##	pro.h:s	#ac.neg	np:p	vother:pred	=rv	%	=other

yup	*=ai.*
sleep	=DECL
v:pred	=rv

'He slept without wearing a blanket.' (observed)

The last example is illustrated by complementation verb serialization, where one serialized verb takes a clause headed by another verb as its complement. Only complement-taking transitive verbs are involved, where the S or A in the complement, which is always coreferential with the S or A argument in the main clause, must not be overtly expressed.

(33)	*shi*	*=gaw*	*ganu*	*=hpe*	#cc	*shat*	*shadu*	%
		3sg	=TOP	mother	=ACC		food	cook
##	pro.h:a	=other	np.h:p	=rn	#cc:p	np:p	vother:pred	%

garum	*=nga*	*=ai.*
help	=CONT	=DECL
v:pred	=aux	=rv

'He is helping his mother cook food.' (elicited)

5. Complex sentences and direct speech

5.1. Subordination and nominalization

The pervasive use and multifunctionality of clausal nominalization are prominent features of Jinghpaw grammar. A nominalized clause can be used not only nominally but also adnominally and adverbially, being exploited to form all the three major types of subordinate clauses: complement clauses, relative clauses (headed and headless), and adverbial clauses. Clausal nominalization is achieved by adding the nominalizer *ai* to a verb, which also marks the verb citation and declarative mood. Thus, an identical clause may occur as a well-formed main clause, complement clause, headed and headless relative clause, and adverbial clause, as illustrated by elicited examples in (34), respectively.

(34) a. *shi*   *shat*   *sha*   *=ai.*
          3sg   food   eat   =DECL
      ## pro.h:a np:p   v:pred   =rv
          'He ate food.' (elicited)

      b. *ngai*   *#cc*   *shi*   *shat*   *sha*   *=ai*   *%*   *chye*
          1sg   3sg   food   eat   =NMLZ   know
      ## pro.h:a #cc:p   pro.h:a np:p   v:pred   =rv   %   v:pred
          =ai.
          =DECL
          =rv
          'I know that he ate food.' (elicited)

      c. *#rc*   *shi*   *shat*   *sha*   *=ai*   *%*   *lakung...*
                  3sg   food   eat   =NMLZ   spoon
      ## #rc pro.h:a np:p   v:pred   =rv   %   np:s
          'The spoon that he ate food with (is)...' (elicited)

      d. *#rc*   *shi*   *shat*   *sha*   *=ai*   *%*   *=hte...*
                       3sg   food   eat   =NMLZ   =COM
      ## #rc:obl pro.h:a np:p   v:pred   =rv   %   =rn
          'With which (spoon) he ate food...' (elicited)

      e. *#ac*   *shi*   *shat*   *sha*   *=ai*   *%*   *majaw...*
                  3sg   food   eat   =NMLZ   because
      ## #ac pro.h:a np:p   v:pred   =rv   %   np:other
          'Because he ate food...' (elicited)

Despite the fact that all the clauses in (34) are headed by the same verb form, we differentiate nominalized (subordinate) clauses from main clauses based on the fact that the former does not exhibit full-fledged properties of main clauses, for example, topic and sentence-final particles never occur within nominalized clauses.

5.2. Complement clauses

Complement clauses, as noted in 5.1, are formed by means of clausal nominalization. The beginning of complement clauses is glossed by the clausal operator ⟨cc⟩, and the end of them by a clause boundary marker ⟨%⟩. Complement clauses may function as the S or P argument, and are thus glossed in the same way as those of other referential expressions. Verbs that have the ability to take nominalized complements may be intransitive verbs from specific semantic classes, such as emotion (e.g., *pyo* 'be fun'), difficulty (e.g., *yak* 'be difficult'), speed (e.g., *lawan* 'be quick'), and judgment (e.g., *teng* 'be true'), or transitive verbs from such semantic classes as knowledge and acquisition of knowledge (e.g., *ce* 'know'), conception (e.g., *shadu* 'think'), perception (e.g., *mu* 'see'), fearing (e.g., *hkrit* 'fear'), preference (e.g., *ra* 'like'), demonstration (e.g., *sharin* 'teach'), manipulation (e.g., *garum* 'help'), and phrasal aspect (e.g., *ngut* 'finish'). Examples:

(35)   a.   #cc   *jawng* *sa*   *=ai*   %   *pyaw*   *=ai*   *=i.*
                             school   go   =NMLZ   be.fun   =DECL   =Q
            ## #cc:s 0.2:s np:g   v:pred   =rv   %   v:pred   =rv   =other
            'Is it fun for you to go to school?' (observed)

       b.   #cc   *marang* *htu*   *=ai*   %   *=gaw*   *ra*
                              rain(n.) rain(v.) =NMLZ   =TOP   like
            ## 0.1:a #cc:p   np:s   v:pred   =rv   %   =other   v:pred
            =ai.
            =DECL
            =rv
            '(I) like rain.' (observed)

Complement clauses, as demonstrated in 4.3, can also be formed by means of verb serialization, in which case, suppression of verbal arguments is observed, unlike nominalized complements, which do not exhibit them. Compare:

(36)   a.   *shi*   *=gaw*   *ganu*   *=hpe* #cc   *shat*   *shadu*
            3sg   =TOP   mother   =ACC   food   cook
            ## pro.h:a =other np.h:p   =rn   #cc:p np:p   vother:pred
       %   *garum*   *=nga*   *=ai.*
            help   =CONT   =DECL
       %   v:pred   =aux   =rv
            'He is helping his mother cook food.' (elicited)

       b.   *shi*   *=gaw*   #cc   *ganu*   *shat*   *shadu*   *=ai*   %
            3sg   =TOP   mother   food   cook   =NMLZ
            ## pro.h:a =other #cc:p   np.h:a   np:p   v:pred   =rv   %
            garum   *=nga*   *=ai.*
            help   =CONT   =DECL
            v:pred   =aux   =rv

'He is helping his mother cook food.' (elicited)

5.3. Relative clauses

Relative clauses, as noted in 5.1, are formed by means of clausal nominalization. Relativization involves no explicit indication of the relationship between the head noun and the relative clause. A relative clause construction may be analyzed as a simple juxtaposition of a nominalized clause and a head noun. This is supported by the flexible position of a relative clause, as shown below, although a relative clause is most commonly prepositive.

(37) a. #rc   *grai*   *gaba*   *=ai*   %   *hpun* *=ni* *moi*
                        very   be.big   =NMLZ   tree   =PL   before
        ## #rc rc_0:s other   v:pred   =rv   %   np:s   =rn   other
        grai   *nga*   *=ai.*
        very   be   =DECL
        other   v:pred   =rv
        'There were many trees which had been very big before.' (elicited)

      b. *hpun* #rc   *grai*   *gaba*   *=ai*   %   *=ni* *moi*
          tree   very   be.big   =NMLZ   =PL   before
        ## np:s #rc rc_0:s other   v:pred   =rv   %   =rn   other
        grai   *nga*   *=ai.*
        very   be   =DECL
        other   v:pred   =rv
        'There were many trees which had been very big before.' (observed)

The "gapped" argument of relative clauses, as in (37), receive the form gloss ⟨rc_0⟩ followed by semantic and function glosses depending on the function of the coreferential head noun, which include not only core arguments such as agent, patient, recipient, and theme, but also obliques, such as companion, instrument, material, vehicle, location, source, goal, and so on. Examples from our corpus include:

(38) a. #rc   *htora*   *lupwa* *=kaw* *yup*   *=ai*   %
                             that   grave   =LOC   sleep   =NMLZ
        ## #rc rc_0.h:s ln_dem   np:l   =rn   v:pred   =rv   %
        dai   *wa*   *=gaw*   *dai*   *=kaw* *yup*   *=na*   *=she...*
        that   man   =TOP   that   =LOC   sleep   =SEQ   =then
        ln_dem   np:s   =other   dem_np:l   =rn   v:pred   =other   =other
        'That man who slept at the grave slept there and...' (KK1-0274_058)

      b. #rc   *ngai*   *=hpe* *lup*   *=da*   *=ai*   %
                                 1sg   =ACC   bury   =RES   =NMLZ
        ##ds #rc 0.h:a rc_0:l pro.1:p   =rn   v:pred =aux   =rv   %

shara	*=kaw*	*nampan*	*langai*	*pu*	*=wa*	*=na*	*re.*
place	=LOC	flower	one	bloom	=VEN	=IRR	COP
np:l	=rn	np:s	rn	v:pred	=aux	=aux	other

'There will bloom a flower at the place where (they) bury me.' (KK1-0474_030)

The head noun is not always coreferential with an argument or adjunct of the modifying clause. In (39a), for example, the modifying clause expresses the content of the head noun, and thus the head noun cannot be interpreted as an argument or adjunct of the modifying clause. Another example comes from (39b), where the head noun, which is not coreferential with an argument or adjunct of the modifying clause, is characterized in relation to the event described by the modifying clause. These examples show that Jinghpaw is a language with a single construction that covers all ranges of the noun modifying clause expressions, which comes to be called the General Noun-Modifying Clause Construction (GNMCC) in the literature (Matsumoto, Comrie, and Sells 2017). These modifying clauses are "gapless," and we assume no gaps (i.e., ⟨rc_0⟩) for these examples.

(39) a.

	bungli	*=hpe*	*=mung*	*atsawm*	*rai*	*galaw*	*=ya*
	work	=ACC	=also	well	LV	do	=BEN
#rc 0.h:a	np:p	=rn	=other	other	other	v:pred	=aux

=mayu	*=ai*	*%*	*myit*	*n-rawng*	*=ai.*
=DESID	=NMLZ		mind	NEG-have	=DECL
=aux	=rv	%	np:s	v:pred	=rv

'They don't have a mind that (they) want to work for (him).' (KK1-0271_014)

b.

	ngai	*=gaw*	#rc	*shu*	*ngoi*	*=ai*	*%*	*nsen*
	1sg	=TOP		frog	make.a.noise	=NMLZ		sound
##	pro.1:a	=other	#rc	np:s	v:pred	=rv	%	np:p

na	*=ai*	*re.*
hear	=DECL	COP
v:pred	=rv	other

'I heard the sound of a frog making a noise.' (observed)

Jinghpaw also has headless relative clauses whose semantic heads are phonologically null. Headless relatives are similar to nominalized complements in that they have a full constituent structure of clauses, and that they constitute an NP head. The empty semantic head of headless relatives may be virtually any semantic role, for example, agent, patient, companion, instrument, location, goal, cause, and so on. Headless relatives that take on argument positions are referential, and thus receive glosses in the same way as those of other referential expressions.

(40) a.

	gaga	#rc	*lusu*	*=ai*	*%*	*=ni*	*=mung*	*sa*	*=ai.*
	other		be.rich	=NMLZ		=PL	=also	go	=DECL
##	ln	#rc.h:s	v:pred	=rv	%	=rn	=other	v:pred	=rv

'At that time, those who are rich also went (to the festival).' (KK1-0187_015)

b.   *nang* *=mung* #rc *nang* *kam* *=ai* %
      2sg   =also   2sg   be.willing =NMLZ
   ##ds pro.2:s =other #rc:l pro.2:s v:pred =rv %

   =kaw *yup* *=u.*
   =LOC sleep =IMP
   =rn   v:pred =rv

   'You also sleep where you like!' (KK1-0274_035)

5.4. Adverbial clauses

Adverbial clauses, except afterthoughts, are preposed to or interposed within main clauses. Jinghpaw has two main strategies to form adverbial subordinate clauses: (a) to employ subordinators that directly follow verbs; and (b) to exploit nominalization-relativization as a subordination strategy with a head noun from generic nouns (e.g., *ten* 'time'), locator nouns (e.g., *hpang* 'after'), and postpositions (e.g., *majaw* 'because').

(41) a. *ganu* *=gaw* #ac *grai* *matsan* *=let* %
     mother =TOP   very be.poor =SIM
    ## np.h:a =other #ac other vother:pred =other %

    =sha *grai* *gasha* *=hpe* *baumaka* *=ai.*
    =only very child =ACC take.care.of =DECL
    =other other np.h:p =rn v:pred   =rv

    'The mother, while being very poor, took very good care of her son.'
    (KK1-0187_005)

  b.   #ac *masusha* *=ai* *majaw* % *grau*
          lie   =NMLZ because   more
    ## 0.h:s #ac 0.h:s v:pred =rv np:other % other

    pawt   *=mayu* *=mat* *=na...*
    get.angry =DESID =COMPL =SEQ
    v:pred =aux =aux   =other

    'Because (they) lied, (he) wanted to get angry much more and...'
    (KK1-0271_061)

5.5. Direct speech

Direct speech (or thought, content, intention, and so on), unlike the subordinate clauses described above, exhibits full properties of sentences, and is thus treated as a full-fledged sentence, not involving nominalization. Direct speech, as illustrated by (42), is introduced by the lexical verb *ngu* 'say that.' This quotative verb, when no addressee is involved, is treated as an intransitive verb as in (42a). It is treated as a transitive verb when, as in (42b), an overtly expressed addressee that is marked by an accusative case just like the P argument function (see 3.3.1) occurs. Direct speech is not analyzed as a P argument in our corpus, but is treated as independent clauses signaled

by a clausal operator ⟨ds⟩.

(42) a.   #ds   *oi*   *sa*   *=ga*   %   *ngu*   *=ai.*
                                     INTJ   go   =HORT   say   =DECL
         ## 0.h:s #ds 0.1:s other v:pred =rv   % v:pred =rv
         '"Hey, let's go," (they) said.' (KK-1861_014)

      b.   *u*   *=ni*   *=wa*   *shi*   *=hpe*   #ds
           bird   =PL   =TOP   3sg   =ACC
         ## np.d:a =rn =other pro.d:p =rn   #ds 0.1:a
         anhte   *=a*   *mun*   *=ni* *shabai la*   *=na*   %   *ngu*
         1pl   =GEN   feather =PL   return take   =IRR   say
         pro.1:poss =rn np:p   =rn   lv   v:pred   =aux   % v:pred
         =nna...
         =SEQ
         =other
         '"We will take back our feather," the birds said to him, and...' (KK-1861_030)

Direct speech is also introduced by means of a quotative complementizer *ngu*. The complementizer, although apparently having a diachronic connection with the quotative verb, is treated as a particle, glossed ⟨other⟩, based on the fact that it is followed by other verbs of utterance and conception, including *ngu* 'say that,' and a full syntactic element may be interposed between complementizers and verbs. Example:

(43)   *madujan* *=gaw*   *maduwa* *=hpe* #ds   *ndai*   *asi*
         wife   =TOP   husband   =ACC   this   fruit
       ## np.h:a =other np.h:p =rn   #ds 0.1:a ln_dem np:p
       =ni *yawnghkra di*   *la*   *=ga*   %   *ngu*   *tsun*   *=ai*
       =PL   all   pick   take   =HORT   QUOT   say   =DECL
       =rn   rn   lv   v:pred   =rv   %   other   v:pred   =rv
       =da.
       =HS
       =other
       'The wife said to her husband, "Let's pick all these fruits," it has been said.'
       (KK1-0269_158)

5.6. Coordination

Jinghpaw does not have a genuine sentence-level coordinating conjunction. A sequence of events is expressed by cosubordination (Foley and Van Valin 1984), where a sequential particle *na* (~ *nna*) is directly added to verbs (or auxiliaries, if any), with only the final verb being marked for aspect and mood. All arguments involving cosubordination can be expressed overtly although often left unexpressed due to their redundancy. A cosubordinate clause, a dependent clause in a strict sense, is simply treated like an independent clause in the Jinghpaw corpus, with its beginning marked

by the leftward-boundary marker ⟨##⟩.

(44) a.   *ngai*   *agatsi*   *=sha*   *lagu* *sa*   *=na*
            1sg   be.silent   =ADV   steal   go   =SEQ
     ##ds pro.1:s other   =other   lv   v:pred =other   ## 0.1:s
     bai   *wa*   *=na*   *=yaw.*
     again   return   =IRR   =SFP
     other   v:pred   =aux   =other
     'I will go silently and secretly and (I) will come back.' (KK1-0263_035)

     b.   *dai*   *=kaw* *rim*   *=na*
                   that   =LOC   catch   =SEQ
     ## 0:a 0.h:p dem_pro:l =rn   v:pred   =other   ## 0:a 0.h:p
     sha   *=kau*   *=ai*   *=da.*
     eat   =away   =DECL   =HS
     eat   =aux   =rv   =other
     '(The tiger) caught (him) there and (it) ate (him), it's said.' (KK1-0265_073)

Coordinating conjunctions are also expressed by means of subordinators *yang*, which form temporal (i.e., 'when') and conditional (i.e., 'if') clauses, as in (45a).[4] This subordinator, often followed by the particle *she* 'then,' is further deprived of its semantic content, as in (45b), being merely used to coordinate successive events. In such case, the clause is often simply treated like an independent clause.

(45) a.   *nanhte*   *ngai*   *=hpe*   *n-mu*   *yang*   *=mung...*
             2pl   1sg   =ACC   NEG-see   when   =also
     #ac.neg pro.2:a pro.1:p =rn   v:pred   other   =other
     'Even if you don't see me...' (KK1-0474_030)

     b.   *bai*   *sa*   *yang*   *=she*   *langai*   *mi*
             again   go   when   =then   one   one
     ## 0.h:s other v:pred other =other ## 0.h:s np.h:l rn
     =kaw *bai*   *du*   *yang*   *=gaw...*
     =LOC   again   arrive   when   =TOP
     =rn   other   v:pred   other   =other
     '(They) went further, and (they) arrived at one person again, and...'
     (KK1-0265_041)

A sequence of events, as shown below, is also encoded by means of serial verb constructions (SVCs). Unlike the abovementioned cases, an SVC is treated as a single clause because serialized verbs form a single predicate. The constraint against role-doubling (Durie 1997), by which a serial verb complex is blocked from containing duplicate roles, that is, two agents, two patients, two instruments, and so on, indicates

---

[4] This kind of neutralization between conditionals and temporal clauses, especially with predictive conditionals and future clauses, is cross-linguistically common since, as Thompson, Longacre, and Hwang (2007: 258) put it, the distinction between temporal and conditional clauses "is simply one of degree of expectability."

that an SVC is monoclausal in contrast to the abovementioned biclausal constructions, which allow duplicate roles to occur within them. For more details of SVCs, see 6.2 below.

(46)   *shi*   =*na*   *manang-wa* =*hpe* *sa*   *shaga* =*ai*
                 3sg   =GEN   friend-man   =ACC   go   call   =DECL
      ## 0.h:a pro.h:poss =rn   np.h:p   =rn   lv   v:pred   =rv
      =*da*.
      =HS
      =other
      '(He) went and called his friend, it's said.' (KK1-0274_060)

6. Constructions with special features

6.1. External possessor constructions

External possessor constructions are constructions where an NP that is semantically understood as the possessor is coded as a core grammatical relation of the verb (Payne and Barshi 1993). In our corpus, external possessors, as exemplified below, are treated as dislocated phrases.

(47) a.   *shi*   =*gaw*   *hkum*   *gaba*   =*ai*.
            3sg   =TOP   body   be.big   =DECL
      ## pro.h:dt =other   np:s   v:pred   =rv
      'As for him, his body is big.' (elicited)

      b.   *shi*   =*gaw*   *kalangta* *nrung* =*mung*   *daw-daw*
            3sg   =TOP   suddenly   horn   =also   be.broken-RED
      ## pro.d:dt =other   other   np:s   =other   v:pred
      re   =*na*   =*i*...
      LV   =SEQ   =SFP
      other =other =other
      'As for her (deer), her horn was also suddenly broken and, OK?'
      (KK1-0263_046)

6.2. Serial verb construction

The pervasive use of serial verb constructions (SVCs), where verbs are serialized productively in a single predicate without any marker of syntactic dependency, is one of the prominent features of Jinghpaw grammar. Serialized verbs are contiguous, and no syntactic elements are interposed between their components. SVCs describe (a) a sequential action, which is expressed by temporally iconic ordering of verbs, where recurrent semantic relationships held between component verbs are consecutivity, means, and cause-effect; and (b) a simultaneously occurring event where component verbs are related in concomitance and manner relationships. Serialized verbs, as a

single predicate, receive only one ⟨v:pred⟩ gloss, which is given to the last verb in serialization. The remaining verbs preceding it automatically receive ⟨lv⟩, regardless of the head of the serialized verbs. As a result of verb serialization, the argument structures of component verbs are conflated into a single structure, following the constraint against role-doubling (5.6). Overt expressions of duplicate roles are systematically suppressed, and thus, no zeros are assumed for them.

(48) a. *dai*   *gwi*   *langai* *mi*   *masha*   *=ni*   *si*   *=mat*
          that   dog   one   one   person   =PL   die   =COMPL
          ## ln_dem np:s   rn   rn   np:a   =rn   v:pred   =aux
          =na   *sa*   *gabai*   *=da*   *=ai*   *=le*   *=i.*
          =SEQ   go   throw.away   =RES   =DECL   =SFP   =SFP
          =other ## 0:p   lv   v:pred   =aux   =rv   =other   =other
          '(The dog) died and the men went and threw it away, OK?' (KK1-0269_098)

      b.   *n-marawn* *shaga*   *=ga.*
                               NEG-shout   speak   =HORT
          ##neg 0.1:s lv   v:pred   =rv
          'Let's not speak by shouting.' (observed)

SVCs, as noted in 5.2, are also exploited for complementation strategies. The complementation serialization is asymmetrical (a term from Aikhenvald 2006) in that the last verbs in the serialization are drawn from a subset of complement-taking verbs, for example, *lanyan* 'be slow,' *ra* 'like,' and *garum* 'help.' SVCs also describe subevents linked by a purposive relationship. In purposive SVCs, as illustrated by (49b), the dependent clause headed by the preceding verb describes the purpose of the following verb in the main clause. Due to the constraint against role-doubling, overt expressions of duplicate roles are systematically suppressed, and, as noted in 4.3, no zeros are assumed for them.

(49) a. *buga-masha* *=ni*   *=gaw*   #cc   *gaga*   *kanbau-bungli*
          local-person   =PL   =TOP   other   living-work
          #ac np.h:dt   =rn   =other   #cc:s   ln   np:p
          lata   % *yak*   *=ai*   *majaw...*
          choose   be.difficult =NMLZ because
          vother:pred   % v:pred   =rv   np:other
          'Because it is difficult for locals to choose other work for a living...' (observed)

      b.   *gwi*   *=hte*   *rau*   #cc   *hpun*   *hta*   %
                      dog   =COM   together   wood   pick
          ## 0.h:s   np:obl   =rn   other   #cc   np:p   vother:pred   %
          sa   *=na*   *=she...*
          go   =SEQ   =then
          v:pred   =other   =other
          '(He) went to gather firewood with his dog and...' (KK1-0269_113)

6.3. Tail-head constructions

Tail-head linkage (THL) is a discourse strategy to connect clause chains by recapitulating the last clause of a chain at the beginning of the next chain (de Vries 2005). THL is also attested in Jinghpaw narrative text. Consider examples (50a) to (50c), which are successively occurring sentences in the same narrative, where every final clause is repeated in the first clause of the next chain to ease processing and/or to carry out discourse-structuring functions, such as referential coherence. The recapitulation, as seen below, is often done partially. We assume zeros in tail clauses in our Jinghpaw corpus.

(50)   a.   *dai*   *magam*   *=gaw*   *num*   *la*   *=s-ai.*
            that   firstborn.son =TOP   woman   take   =CSM-DECL
            ## ln_dem np.h:a   =other   np.h:p   v:pred   =rv
            'The firstborn son married a wife.' (KK1-0262_032)

       b.   *num*   *la*   *=na*   *=she*   *nga*
                       woman   take   =SEQ   =then   live
            ## 0.h:a np.h:p   v:pred   =other   =other   ## 0.h:s   v:pred
            rai   *yang*   *=she*   *dai*   #rc_0:p   *shan*
            LV   when   =then   that   3du
            other   other   =other   ## ln_dem #rc_0:p   pro.h:a
            nau   *hkai*   *=da*   *=ai*   %   *hpun*   *dai*   *=mung*
            brother   plant   =RES   =NMLZ   tree   that   =also
            np.h:appos   v:pred   =aux   =rv   %   np:dt_s   dem_np:s   =other
            gaba   *=wa*   *=s-ai.*
            be.big   =VEN   =CSM-DECL
            v:pred   =aux   =rv
            ' (He) married a wife and (they) lived and the tree that the brothers planted also grew.' (KK1-0262_033, 034)

       c.   *gaba*   *=wa*   *=na*   *=she...*
                      be.big   =VEN   =SEQ   =then
            ## 0.h:s v:pred   =aux   =other   =other
            '(The tree) grew and...' (KK1-0262_035)

6.4. Phrase or clause repetition

Repetition of phrases and clauses prevails in Jinghpaw narrative texts. It, as illustrated by (51a), contributes to an iconic meaning associated with concepts such as iterativity and durativity. Repetition also performs the function of reinforcing communication, as in (51b), where the speaker repeats the command to ensure the hearer's attention. Repeated constructions are counted only once following Bickel

(2003), and leaving others glossed ⟨nc⟩ (i.e., 'non-classifiable.')[5]

(51) a.   *hkawm hkawm hkawm re   =she...*
               walk   walk   walk   LV   =then
         ## 0.h:s nc   nc   v:pred other =other
         'He walked, walked, walked, on and on...' (KK1-0269_158)

      b.   *mare   dai   =kaw   =wa   =she   masha   kadai n-nga*
                village that =LOC   =TOP   =then   person   who   NEG-live
         ##neg   np:l   rn   =rn   =other =other   np.h:s   rn   v:pred
         =taw-nga   =ai   =da.   masha   kadai =mung n-nga
         =CONT-CONT =DECL =HS   person   who   =also NEG-live
         =aux   =rv   =other ##nc   nc   nc   nc   nc
         =taw-nga   =ai.
         =CONT-CONT =DECL
         nc   nc
         'No one lived in the village, it's said. No one lived in the village.'
         (KK1-0274_034, KK1-0274_035)

7. Initial findings

This section presents our interim findings based on 1,221 annotated clauses by comparing them with findings of previous studies (Bickel 2003, Du Bois 2003, Noonan 2003, Haig, Schnell, and Wegener 2011, Haig and Schnell 2016a, Brickell and Schnell 2017, among others). The raw data from our interim corpus, from which our initial findings stem, are given below:

Table 3   Raw data (third person only)

	lexical		pronoun		zero		total
	[+hum]	[−hum]	[+hum]	[−hum]	[+hum]	[−hum]	
S	117	111	106	10	279	50	673
A	70	4	77	0	165	1	317
P	61	189	24	12	48	94	428
total	248	304	207	22	492	145	1418

---

[5] The gloss ⟨nc⟩ is also given to phrases and sentences that are not inside the storylines, for example, titles of the stories.

Table 4   Raw data (all persons)

	lexical		pronoun		zero		total
	[+hum]	[−hum]	[+hum]	[−hum]	[+hum]	[−hum]	
S	117	111	151	10	325	50	764
A	71	4	150	0	248	1	474
P	61	189	48	12	63	94	467
total	249	304	349	22	636	145	1705

7.1. Core arguments and human expressions

Haig, Schnell, and Wegener (2011), based on four GRAID-annotated text corpora of Awetí (Tupi-Guarani), Gorani (Indo-European, Iranian), Savosavo (Papuan Isolate), and Vera'a (Austronesian, Oceanic), present analyses on the distribution of forms of core arguments, human expressions, and associated phenomena based on the proportion of (a) S, A, and P arguments, (b) intransitive and transitive clauses, (c) human vs. non-human arguments in core arguments, and (d) human arguments among S, A, and P arguments. This section, along the same vein, provides related figures from our Jinghpaw corpus and some remarks on them.

The proportion of S, A, and P arguments drawn from the Jinghpaw corpus is given in (52).[6]

(52)   Proportion of S, A, and P arguments (all persons)
    S   44.8% (764)
    A   27.8% (474)
    P   27.4% (467)

Related to this is the proportion of intransitive and transitive clauses given in (53), which is calculated by taking the overall P arguments as representative of transitive clauses. What we have here is striking in that the ratio of intransitive to transitive clauses are roughly two thirds to one third, replicating the same proportion obtained by Haig, Schnell, and Wegener (2011: 68) in other languages.

(53)   Proportion of intransitive and transitive clauses (all persons)
    intransitive   62.1% (764)
    transitive   37.9% (467)
    total   100% (1,231)

The proportion of human vs. non-human arguments in core arguments is given in (54) below, where non-human arguments involve both human and anthropomorphized discourse participants. Again, what is of importance here is the fact that the proportion

---

[6] The A and P arguments, as a rule, should show the same proportions because they always co-occur in a transitive clause. The slightly higher proportion of A arguments should be accounted for in further investigation.

where three-fourth of the core arguments are human participants is roughly consistent with the proportion (two-thirds) obtained by Haig, Schnell, and Wegener (2011: 69) based on other languages.

(54)   Proportion of human vs. non-human arguments in core arguments (all persons)
       [+hum]   72.4% (1,234)
       [−hum]   27.6% (471)

The proportion of human arguments among S, A, and P arguments is given in (55), where human referents are often expressed as S or A in contrast to P, which often involves non-humans. This asymmetrical proportion is consistent with the crosslinguistically common tendency where the information flows from A to P and from more animate to less animate, as Comrie (1981: 121) puts it: "in actual discourse there is a strong tendency for the information flow from A to P to correlate with an information flow from more to less animate and from more to less definite. In other words, the most natural kind of transitive construction is one where the A is high in animacy and definiteness, and the P is lower in animacy and definiteness; and any deviation from this pattern leads to a more marked construction."

(55)   Proportion of human arguments among S, A, and P arguments (all persons)
       S [+hum]   77.6% (593)
       S [−hum]   22.4% (171)
       A [+hum]   98.9% (469)
       A [−hum]   1.1% (5)
       P [+hum]   36.8% (172)
       P [−hum]   63.2% (295)

This tendency, as noted in 3.3.1, motivates the alignment of core case marking where P is obligatorily case-marked when it outranks or is equal to A on the animacy hierarchy because this situation deviates from the more general pattern of the information flow.

7.2. Preferred Argument Structure and the discourse basis of ergativity

Du Bois (1987, 2003) suggests that the argument realization in discourse is systematically shaped by the violable soft constraints given in (56), where (56a) and (56b) are quantity-related and (56c) and (56d) are role-related constraints. Together these form Preferred Argument Structure (PAS), which is suggested to be a discourse universal.

(56)   Preferred Argument Structure constraints
         a. Avoid more than one lexical core argument per clause.

         b. Avoid more than one new core argument per clause.

         c. Avoid lexical A.

         d. Avoid new A.

The figure in (57) shows the proportion of lexical expressions within each of S, A, and P in our Jinghpaw corpus, answering the question "How lexical is each role?" (Haig and Schnell 2016b: s6–9). It shows that the proportion of lexical A (23.3%) is low compared to S (33.9%) and P (58.4%), confirming the nonlexical A constraint (56c). It should be noted, however, that the proportion of lexical A in the Jinghpaw corpus is still high when compared to major findings obtained by other studies (Du Bois 2003: 37, Haig and Schnell 2016a: 599, Brickell and Schnell 2017: 197). Note that our counts, following Haig and Schnell (2016a) and Brickell and Schnell (2017), exclude first and second persons, the inclusion of which decreases the overall lexicality of A. Further studies based on more data are required to understand how the Jinghpaw pattern arises.

(57)   Proportion of lexical expressions within each of A, S, and P (third person only)
         A   23.3% (74 of total of 317 A arguments are lexical)
         S   33.9% (228 of total of 673 S arguments are lexical)
         P   58.4% (250 of total of 428 P arguments are lexical)

Recent studies show that the low lexicality of A is an epiphenomenal by-product of two more general aspects of humanness and topicality (Everett 2009, Haig and Schnell 2016a, Brickell and Schnell 2017), as Brickell and Schnell (2017: 204) put it: "patterns of argument realisation in discourse are epiphenomenal of humanness and – since human referents are usually discourse-topical in narratives – topicality. The rationale here is that topical referents are often realised non-lexically, and thus human referents frequently take non-lexical form. Since humans are more likely to fulfil agent-like roles and non-human referents are more likely to have patient-like roles, A's are least and P's most likely to be realised lexically." This association between humanness and agentivity is also replicated by our Jinghpaw data, as illustrated by (58), which shows a significantly high proportion of [+hum] arguments in A function (including both human and anthropomorphized discourse participants) regardless of their formal encoding (lexical nouns, pronouns, and zeros).

(58)   Proportion of [+hum] arguments in A function (third person only)
         [+hum] A   98.4% (312)
         [−hum] A   1.6% (5)

Note that all examples of [−hum] A in our data come from non-anthropomorphized animal A arguments, as illustrated by (59). This fact shows that inanimate A arguments are significantly dispreferred in Jinghpaw discourse.

(59)   Examples of [−hum] A
         'this *horse* ate all our paddy seeds...' (KK1-0271_057)
         'This *dog* does not know a human language...' (KK1-0269_119)
         'The evidence about what food this *cow* ate is here...' (KK1-0272_076)

The figure in (60) shows the lexicality of human vs. nonhuman A in Jinghpaw corpus.

The fact that the proportion of nonhuman lexical A's (80%) and that of human lexical A's (22.4%) significantly differ from each other supports the position that the source of the low lexicality of A is humanness rather than the A role itself, as demonstrated by Haig and Schnell (2016a: 609–12).

(60) Lexicality of [+hum] vs. [−hum] A (third person only)

	[+lex]	[−lex]	Totals
[+hum]	70	242	312
[−hum]	4	1	5
Totals	74	243	317

In addition to the Preferred Argument Structure constraints given in (56), Du Bois (1987, 2003) claims the unity of S and P in opposition to A based on the equally high lexicality of S and P as opposed to the low lexicality of A. This constitutes the "discourse basis of ergativity." More recent studies, however, have shown that the ergative alignment in discourse is not supported by empirical data from many languages (Kumagai 2006, Everett 2009, Haig and Schnell 2016, Brickell and Schnell 2017). The same holds for our Jinghpaw data, where no significant clustering of S and P is observed, as repeated here as (61) for easy reference.

(61) Proportion of lexical expressions within each of A, S, and P (third person only)
    A   23.3%
    S   33.9%
    P   58.4%

Our data show that, in terms of lexicality, S is located between A and P, which is consistent with observations by Haig and Schnell (2016a). They further show a split of S, where the nonhuman S tends toward P in contrast to the human S, which tends toward A. Rather than assuming the nonlexical A constraint, they formulate a more general tendency given in (62), which explains the impact of humanness.

(62) Haig and Schnell (2016a: 612)
    A and S, if they refer to human referents, are seldom lexical.

Our data drawn from Jinghpaw, as given in (63), also show the split of S, where the human S tends toward A as opposed to the nonhuman S, which tends toward P, confirming thus the general tendency (62).

(63) Proportion of lexical expressions within each of A, S [+hum], S [−hum], and P
    (third person only)
    A   23.3%
    S [+hum]   23.3% (117 [+lex, +hum] S of total of 502 [+hum] S arguments)
    S [−hum]   64.9% (111 [+lex, −hum] S of total of 171 [−hum] S arguments)
    P   58.4%

7.3. Referential density

Referential density (RD) is the ratio of overtly expressed arguments to available argument positions in the clause (Bickel 2003, Noonan 2003). The lower the referential density is, the more zero-anaphora appears in the discourse. Bickel (2003), based on Pear Story narratives of three languages in the Nepalese Himalayas, that is, Belhare (Tibeto-Burman), Maithili (Indo-Aryan), and Nepali (Indo-Aryan), shows that RD varies significantly from language to language despite the fact that speakers of these languages share a common socio-cultural setting. Below is the ranking of RD values given by Noonan (2003) based on Frog Story narratives or other alternatives from several languages.

(64)   (somewhat arbitrary) ranking of RD values (Noonan 2003: 6)
      < 0.50   = low   (e.g., Belhare)
      0.50-0.70 = moderate (e.g., Japanese)
      0.70 >   = high   (e.g., English)

The RD value in our data, although not straightforwardly compatible with more controlled data given by Bickel (2003) and Noonan (2003) because of the difference in genre, is given in (65), where (65a) is based on 781 overtly expressed arguments (both lexical nouns and pronouns) within 1,418 available argument positions in the clause, and (65b) is based on 924 overtly expressed arguments within 1,705 available argument positions in the clause. The comparison between (65a) and (65b) shows that the RD value is relatively stable irrespective of person types involved.

(65)   The RD value in the Jinghpaw texts
      a. 781/1418=0.55 (excl. first and second persons)
      b. 924/1705=0.54 (incl. first and second persons)

The RD values indicate that our corpus is rife with zero-anaphoras, where approximately half of argumental functions are not overtly expressed. This is not surprising given that Jinghpaw NPs are freely omitted when they are pragmatically retrievable from the context (3.1.3). Jinghpaw discourse thus requires listeners to infer much about referents, as illustrated by examples given in 3.1.3 above.

8. Conclusions

This paper explored an implementation of the GRAID glossing conventions to a corpus of Jinghpaw, explaining structures of major clauses, annotations of referential expressions in terms of forms (NPs, pronouns, zeros), semantics (person and animacy), and functions (S, A, P, obliques, etc.), and types of predicates (verbal, nominal, copula, and non-canonical predicates), and annotations of complex sentences, including complements, relative clauses, adverbial clauses, direct speech, and sentence-level

coordination. These are followed by initial findings drawn from the annotated Jinghpaw corpus, especially focusing on Preferred Argument Structure, the discourse basis of ergativity, and referential density. Our findings drawn from Jinghpaw showed (a) that the low lexicality of A can be replicated by our data; (b) that the ergative alignment in discourse is not supported by our data; and (c) that approximately half of argumental functions are not overtly expressed in our texts. Our Jinghpaw data also confirmed other crosslinguistic discourse tendencies, including the avoidance of inanimate A, and the information flow from A to P and from animate to inanimate.

Symbols and abbreviations

- affix boundary
= clitic boundary
1 first person
2 second person
3 third person
sg singular
du dual
pl plural
ABL ablative
ACC accusative
ADV adverbializer
ALL allative
BEN benefactive
COM comitative
COMPL completive
CONT continuous
COP copula verb
CSM change-of-state marker
DECL declarative
DESID desiderative
GEN genitive

HORT hortative
HS hearsay
IMP imperative
INTJ interjection
IRR irrealis
LOC locative
LV light verb
NEG negative
NMLZ nominalizer
PL pluralizing clitic
Q question
QUOT quotative
RED reduplicant
RES resultative
SEQ sequential
SFP sentence-final particle
SIM simultaneous
TERM terminative
TOP topic
VEN venitive

Acknowledgements

References

Aikhenvald, Alexandra Y. 2006. "Serial verbs constructions in a typological perspective". In Alexandra Y. Aikhenvald and R.M.W. Dixon (eds.) *Serial verb constructions: A cross-linguistic perspective*. Oxford: Oxford University Press. pp.1–68.

Bickel, Balthasar. 2003. "Referential density in discourse and syntactic typology". *Language* 79(4). pp.708–736.

Brickell, Timothy C. and Stefan Schnell. 2017. "Do grammatical relations reflect information status? Reassessing Preferred Argument Structure theory against discourse data from Tondano". *Linguistic Typology* 21(1). pp.177–208.

Comrie, Bernard. 1981. *Language universals and linguistic typology*. London: Basil Blackwell.

Dixon, R. M. W. 1977. "Where have all the adjectives gone?" *Studies in Language* 1. pp.19–80.

Du Bois, John W. 1987. "The discourse basis of ergativity". *Language* 63(4). pp.805–855.

———. 2003. "Argument structure: Grammar in use". In John W. Du Bois, Lorraine E. Kumpf, and William J. Ashby (eds.) *Preferred Argument Structure: Grammar as architecture for function*. Amsterdam: John Benjamins. pp.11–60.

Durie, Mark. 1997. "Grammatical structures in verb serialisation". In Alex Alsina, Joan Bresnan, and Peter Sells (eds.) *Complex Predicates*. Stanford: CSLI. pp.289–354.

Everett, Caleb. 2009. "A reconsideration of the motivations for preferred argument structure". *Studies in Language* 33(1). pp.1–24.

Foley, William and Robert Van Valin, Jr. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Haig, Geoffrey and Stefan Schnell. 2014. "Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators (version 7.0)". In Geoffrey Haig and Stefan Schnell (eds.) *Multi-CAST (Multilingual Corpus of Annotated Spoken Texts)*. https://lac.uni-koeln.de/en/multicast/ (accessed 2017-10-14.)

———. 2016a. "The discourse basis of ergativity revisited". *Language* 92(3). pp.591–618.

———. 2016b. "The discourse basis of ergativity revisited: Online appendices". *Language* 92(3). pp.s1–s14.

Haig, Geoffrey, Stefan Schnell, and Claudia Wegener. 2011. "Comparing corpora from endangered language projects: Explorations in language typology based on original texts". In Geoffrey Haig, Nicole Nau, Stefan Schnell, and Claudia Wegener (eds.) *Documenting endangered languages: Achievements and perspectives*. Berlin: Mouton de Gruyter. pp.55–86.

Kumagai, Yoshiharu. 2006. "Information management in intransitive subjects: Some implications for the preferred argument structure theory". *Journal of Pragmatics* 38. pp.670–694.

Kurabe, Keita. 2012. "Jinpōgo no kakuhyōji [Case marking in Jinghpaw]". *Kyoto University Linguistic Research* 31. pp.113–180.

———. 2017. Recordings of Jinghpaw folktales (KK1). Digital collection managed by PARADISEC. [Open Access] DOI: 10.4225/72/59888e8ab2122.

LaPolla, Randy J. 1992. "Anti-ergative marking in Tibeto-Burman". *Linguistics of the Tibeto-Burman Area* 15(1). pp.1–9.

Malchukov, Andrej. 2008. "Animacy and asymmetries in differential case marking". *Lingua* 118. pp.203–221.

Matsumoto, Yoshiko, Bernard Comrie, and Peter Sells. 2017. *Noun-modifying clause constructions in languages of Eurasia: Rethinking theoretical and geographical boundaries*. Amsterdam and Philadelphia: John Benjamins.

Noonan, Michael. 2003. "A cross-linguistic investigation of referential density". Online publication, available at http://archiv.ub.uni-heidelberg.de/savifadok/volltexte/2008/190/.

Payne, Doris L. and Immanuel Barshi (eds.) 1999. *External possession*. Amsterdam: Benjamins.

Schnell, Stefan and Nils N Schiborr. 2018. "Corpus-based typological research in discourse and grammar GRAID and Multi-CAST". *Asian and African Languages and Linguistics* 12. pp.1–16.

Thompson, Sandra A., Robert Longacre, and Shin Ja J. Hwang. 2007. "Adverbial clauses". In Timothy Shopen (ed.) *Language typology and syntactic structure*, vol. 2. Cambridge: Cambridge University Press. pp.237–300.

de Vries, Lourens. 2005. "Towards a typology of tail-head linkage in Papuan languages". *Studies in Language* 29(2). pp.363–384.