Panel Session: Tibeto-Burman languages of the Indo-Myanmar borderland

Scott DeLancey
University of Oregon
delancey@uoregon.edu

The languages of Manipur, Nagaland, and the far eastern reaches of Assam and Arunachal have been very little studied until very recently. For the several dozen languages of this area we have less than a handful of detailed descriptions (Chelliah 1997, Coupe 2007) and a few sketch grammars and unpublished dissertations at local universities (e.g. Kongkham 2010). This panel will report ongoing research on three very poorly-understood groups of Tibeto-Burman languages of the Indo-Myanmar border, and summarize some results for both Tibeto-Burman and broader typology which have or promise to emerge from this research. The lead presenter will be Scott DeLancey, University of Oregon; the rest of the panel will be three current or recent Ph.D. students who are conducting field research in the region. Dr. Linda Konnerth on the Northwest Kuki-Chin languages of Manipur, particularly Monsang; Krishna Boro on the Tangsa-Nocte languages of upper Assam, particularly Hakhun Tangsa; and Amos Teo on Angami-Pochuri languages of Nagaland, particularly Sumi.

The Nocte, Tangsa, and Tutsa languages in the Northern Naga branch, and the Northern (NKC) and Northwest or "Old Kuki" (NWKC) branches of Kuki-Chin in Manipur and Chin State, index the arguments of the verb by means of AGREEMENT WORDS, person-number indices which are phonologically independent rather than suffixed to the verb stem (first noted by Henderson (1965) and Stern (1963)):

Nocte (N Naga)	1)	*ŋaa*	*ka*	*ʌ̀ŋ*
		I	go	1SG
		'I go.'		
	2)	*nʌ̀ŋ*	*ka*	*ɔ*
		you	go	2SG
		'You.SG go.'		
Tedim (N KC)	3)	*pài*	*ìŋ*	
		go	1SG	
		'I go.'		
	4)	*pài*	*tɛ̲ʔ*	
		go	2	
		'You.SG go'		

These may be simple agreement indices, or combine with tense/aspect or negation:

Hakhun (N Naga)	5)	*ŋa*	*TV*	*su* *k-ɤ*
		I	TV	watch PRES-1SG
		'I am watching TV'		
Monsang (NW KC)	6)	*kɤ*	*i-si*	*k-iŋ*
		I	NMZ-go	PRES-1SG
		'I am going/ will go.'		

These constructions originated in inflected auxiliary verbs – the simple agreement words in exx. 1-4 reflect old inflected copulas – but synchronically they are phonologically independent elements of the verb complex, corresponding to grammaticalized suffixes in other TB languages (DeLancey to appear a, under review). This phenomenon is no longer found in southern Northern Naga languages, which the agreement words have simply been dropped, or in the Central Kuki-Chin languages, where the old conjugation has been completely replaced by a new prefixal agreement system (DeLancey 2013a).

These languages also show hierarchical agreement and inverse marking. Both phenomena can be reconstructed for Proto-Tibeto-Burman, but in both Nocte-Tangsa and NWKC the original system has been lost, and we find innovative secondary inverse constructions. Hakhun (N Naga) has a very new inverse construction derived from a motional cislocative (Boro 2012, cp. DeLancey 2011):

O / A	1SG	2SG	3SG
1SG		Σ ɤʔ	Σ ɤ
2SG	Σ r-ɤ		Σ oʔ
3SG		Σ r-u	Σ a

Table 1: Inverse *r-* in the Hakhun progressive conjugation (Nocte-Tangsa)

The Monsang (NWKC) inverse derives from an impersonal construction (Konnerth and Wanglar 2014):

O / A	1SG	2SG	3SG
1SG		ki-Σ na-tsə	ki-Σ naʔ
2SG	m-Σ na-tsə		na-Σ naʔ
3SG	m-Σ naʔ		**a**-Σ naʔ

Table 2: Inverse *m-* in the Monsang progressive conjugation (NW Kuki-Chin)

The *m-* is cognate to the 1st object index *mi-* in Mizo, both < **mi* 'person'. Pending the availability of more comparative NW and Northern KC data it is not clear whether the inverse function found in Monsang is recent or of Proto-NWKC or possibly Proto-KC provenance.

Thus the languages at the northern and southern ends of the Purvanchal Range show very similar grammatical structures, incorporating both ancient morphological material and innovative patterns which recreate older pan-TB structures such as inverse marking. In the 200-kilometer belt of "Naga" languages from mid-Manipur through Nagaland these structures are absent, and this shared grammatical profile was one reason that these languages were previously thought to form a genetic unit. This hypothesis has been universally abandoned over the last few years (Burling 2003); it is now becoming clear that the shared grammatical tendencies of the "Naga" group represent horizontal rather than vertical transmission.

Some "Naga" languages show incipient re-complexification; for example, Sumi has innovated new 1st and 2nd person Object agreement, based on possessive proclitics (Teo 2014):

Sumi	7)	*nò=nò*	*ì=hè*
		2SG=AGT	1SG=hit
		'You hit me.'	
	8)	*ì=nò*	*ò=hè*
		1SG=AGT	2SG=hit
		'I hit you.'	

But in general the languages of Nagaland and northern Manipur show the more shallowly-grammaticalized "creoloid" grammatical structure characteristic of TB languages which have been subject to intense language contact (DeLancey 2013b).

The languages under consideration offer much of interest to general and areal typology and to the classification and reconstruction of Tibeto-Burman. Inverse systems and hierarchical agreement are topics of current theoretical and typological interest; the Nocte-Tangsa and NWKC languages present new instances of these patterns and new perspectives on their historical development. The agreement forms in these languages represent inheritance ultimately from Proto-Tibeto-Burman (DeLancey 2011, 2013a, c, 2014a), probably with a more recent shared history as well (DeLancey under review). They thus offer new data for reconstructing the original patterns, and examples illustrating the development and reorganization of complex indexation structures. The inverse systems in particular are clearly new, so these languages constitute a laboratory in which we can see the recent development of these phenomena.

The typological split between the complex structures in the northern and southern languages and the creoloid patterns in the Naga Hills crosses genetic lines, certainly in the north, and probably in the south. In the north, the Northern Naga group, which certainly represents a genetic unit, is split between the archaic Nocte, Tutsa and Tangsa languages, and, just within the Naga area, languages like Wancho, Konyak, Chang and others which have lost the agreement word system. There appears to be a similar situation to the south of the Naga area; Mortensen and Keogh (2011) argue that Sorbung, and likely other languages of southern Ukhrul District in Manipur, are NWKC in origin, but have adapted to Naga syntactic patterns. So it appears that the explanation for the typological variation up and down the Purvanchal Range lies in local areal typological developments. This dramatic difference between the "Naga" center and the two peripheries is of great potential interest for the study of "complexity" (Dahl 2004, Trudgill 2011, DeLancey 2014b).

In the panel presentation, Boro, Konnerth and Teo will each present a quick typological overview of the languages they are working on, with special attention to some of the features discussed here. DeLancey will present an overview of the typological and historical implications of these data, and the areas which at present cry out for further research on the ground.

References

Boro, Krishna. 2012. "Verb agreement in Hakhun Tangsa." Paper presented at the 18th Himalayan Languages Symposium, Varanasi.

Burling, Robbins. 2003. "The Tibeto-Burman languages of Northeastern India". In: Thurgood, Graham, and Randy LaPolla (eds.), *The Sino-Tibetan Languages*. London: Routledge, 169-191.

Chelliah, Shobhana. 1997. *A Grammar of Meithei*. Berlin: Mouton de Gruyter.

Coupe, Alexander. 2007. *A grammar of Mongsen Ao*. Berlin; New York: Mouton de Gruyter.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam & Philadelphia: Benjamins.

DeLancey, Scott. 2011. "Nocte and Jinghpaw: Morphological Correspondences". In: Hyslop et al. (eds.), *North East Indian Linguistics III*.

_____. 2013a. "Argument indexation (verb agreement) in Kuki-Chin". 46th International Conference on Sino-Tibetan Languages and Linguistics.

_____. 2013b. "Creolization in the divergence of Tibeto-Burman". In Owen-Smith and Hill, (eds.), *Trans-Himalayan Linguistics: Historical and Descriptive Linguistics of the Himalayan Area*.

_____. 2013c. "The history of postverbal agreement in Kuki-Chin". *Journal of the Southeast Asian Linguistics Society* 6:1-17.

_____. 2014b. "Sociolinguistic typology in Northeast

India: A tale of two branches." *Journal of South Asian Languages and Linguistics* 1.1: 59-82.

_____. 2014a. "Second person verb forms in Tibeto-Burman." *LTBA* 37.1: 3-33.

_____. to appear a. "Prosodic typology and the morphological structure of the verb in Tibeto-Burman". *Nepalese Linguistics*.

_____. to appear b. "Deictic and sociopragmatic effects in Tibeto-Burman SAP indexation."

_____. under review. "Morphological evidence for a Central branch of Trans-Himalayan (Sino-Tibetan)."

Henderson, E.J.A. 1965. *Tiddim Chin: A descriptive analysis of two texts*.

Kongkham, Hemabati. 2010. *A Descriptive Grammar of Moyon*. Ph.D. dissertation, Department of Linguistics, Manipur University.

Konnerth, Linda, and Koninglee Wanglar. 2014. Monsang, a Northwestern Kuki-Chin language from Manipur. 47th International Conference on Sino-Tibetan Languages and Linguistics.

"Graveyard of Languages" (?): A Study of Nine Critically Endangered Languages of North Bengal

Indranil Acharya
Vidyasagar University, West Bengal
acharya.indranil@gmail.com

Atasi Nanda Goswami
ICSSR, New Delhi
Ektuatasi@gmail.com

Susil Mandal
PLSI, West Bengal
susilmondal1959@gmail.com

1 Linguistic Survey, Census and the Facts

After the completion of linguistic survey by George Abraham Grierson during 1894-1927 the total number of languages reported was 733. In the 1961 census the language count was 1652 in India. But the 2001 census recorded a sharp decline in the number of languages. It was reported to be only 122. The most obvious factors responsible for this major slump were the government policy of recognizing only those languages that have a speaker base of more than 10,000 and the ever increasing importance of Hindi and English as killer languages at the national level. But on a deeper probe the main reason of the death of so many languages turns out to be the absence of livelihood options for the speech communities in their own languages. Another genuine factor for this endangerment has been a highly lopsided development that throttled the voices of smaller languages. Various nomadic communities are fast losing their languages as they feel insecure to disclose their linguistic identity. Other major victims of developmental initiatives are the coastal speech communities who had to migrate inland due to drastic changes in sea-farming laws (Devy 2013).

2 Language Endangerment: Facts and Factors

According to G.N.Devy (Devy 2013), the worldwide phenomenon of language endangerment today is the direct offshoot of a fundamental shift in production practices. This crisis occurred before when hunting-gathering society shifted to agriculture (about 7 to 10 millennia before human civilization). At that point of time natural languages died and newer languages with a different sense of grammar emerged. This shift in production practices has severely affected the existence of hundreds of small languages. As many as 310 languages- 263 spoken

by less than 05 people and 47 by less than 1,000- are nearing extinction. These 310 languages were living during the 1961 census. Only 10 languages barely survive at present. It is disheartening to learn that barely one-fifth of India's linguistic heritage has been extinct over the last fifty years (Bandyopadhyay 2013). Rapid industrialization has forced large scale 'language migrations' and the smaller speech communities have been coerced to speak other languages of greater political significance. Livelihood options in their homeland are also dwindling at an astonishing rate. Keeping in mind the rich linguistic diversity of India (it is ranked 9th among the top linguistically diverse countries with a score of .930), it is a profoundly disturbing situation and needs urgent public attention (Dasgupta 2013).

3 Endangered Languages of North Bengal: Status Report

Against this dismal picture of severe endangerment of languages the nine critically endangered languages of North Bengal have to be studied. The languages are- Tharu, Hajong, Jalda, Kaya, Dhuliya, Rautiya, Shabar, Asur and Magar. The 10th language- Tundu- is totally extinct by now. In the Mal subdivision of the Jalpaiguri, there is a tea garden called Tundu. The early inhabitants of this region were the people of the Tundu tribe. With the passage of time, they have been absorbed into the Oraon Malpahariya community. One finds no evidence today of their existence as a separate tribe. The Tundus are commemorated only by the eponym of the place. The Tharus mainly occupy the region of Naxalbari, Fakirchan Jot, Ghoshpukur and Tharu Jot in the Terai regions of Darjeeling. They have begun to forget the heritage of their mother-tongue and to align themselves with the Hindu Rajbangshi linguistic traditions. Although a few members of the Hajong tribe also live in the Cooch Behar and Jalpaiguri districts, they have gradually been absorbed into the larger Rajbangshi society. Thus, their language, too, is endangered. Jaldapara in the Dooars used to be the native land of the Jaldas. They have today lost touch with their mother tongue as a result of being absorbed into the larger Rajbangshi culture. At present, there are a few speakers of Kaya in the Dooars of North Bengal. They still preserve their own culture, rituals, and way of life. However, they have gradually become aligned with the Oraon community, and the language is beginning to die out. The Dhuliya tribe belongs to the north of the Kamakhyaguri region of Jalpaiguri. Since the last decade of the 19th century, in the midst of the Kshatriya movement that arose among the Rajbangshis, they have mingled with the larger Rajbangshi community. The language of the Rautiyas is endangered today. Most Rautiyas do not know their mother-tongue. They have adopted Sadri as their language for use at home and outside. The Shabars inhabit the tea-estates of the Terai and Dooars in North Bengal. From what can be gathered of their origins, they migrated to the tea-gardens in North Bengal from the Ganjam, Ajaygarh, Parlaand Kemonti regions of Orissa. Although the Shabars speak their mother tongue at home, they speak Sadri, Hindi and Bangla outside. Thus, their language is endangered. The Asur tribe inhabits the Karon, Kurti and Dalsingpara tea gardens of the Dooars and Majher Dabri in the Alipurduar block. Their language is threatened, and they are increasingly beginning to use Sadri as the language of communication both at home and outside. As members of the Nri family, the Magars are affiliated to the Mongolian group. They are giving up their mother-tongue and instead adopting other languages like Nepali and Sadri.

4 Research Findings: Grammar and Vocabulary

Intensive research on these endangered languages was initiated and sustained by the People's Linguistic Survey of India over the last couple of years. To recognize a 'language' as such the field linguists and investigators set up two major criteria-

- to check whether the language has its own set of grammatical rules and
- to verify whether the language possesses a unique vocabulary up to 70%.

Going by these criteria it was found that out of the 09 endangered languages only 04- Hajong, Shabar, Asur and Magar- have distinct vocabulary and certain recognizable grammatical features- phonological, morphological and syntactical. In the remaining 05 languages vocabulary and grammatical structure were difficult to retrieve. The languages had fallen into disuse to a large extent. There is a huge language gap between the older generation (60-80 years) and the younger generation (10-30 years) and it has widened at an alarming rate in recent years. A 20-year-old young man from Dhuliya community cannot properly string a single sentence in his mother tongue. This is a major factor responsible for endangerment.

5 Conclusion: Need for Further Research

A landmark research work in the domain of endangered languages was done by Anvita Abbi (Abbi 2006) when she focused on the Bo language

of the Great Andamanese family of languages. When the last speaker Boa Senior died at the age of 85 in January 2010, the linguistic identity of the tribe disappeared. Anvita Abbi did a great job in representing a dying linguistic community with great research zeal. Unfortunately, the languages of North Bengal are yet to receive such scholarly attention. In the final analysis one needs to re-think and re-formulate the research strategies for the preservation of these near extinct languages.

6 Acknowledgements

Special mention should be made of Bhasha, Baroda for undertaking People's Linguistic Survey of India that helped to identify these critically endangered languages. The findings will be included in the volume on the Languages of West Bengal to be published by Orient Blackswan shortly.

References

Abbi, A. 2006. *Endangered Languages of the Andaman Islands.* Muenchen: Lincom Europa.

Bandyopadhyay, K. 2013. "10 languages in state 'critically endangered'". *The Times of India* August 31: 6.

Dasgupta, D. 2013. "Speaking of Us". *Outlook* September 09: 5-11.

Devy, G.N. 2013. *The Being of Bhasha: A General Introduction*, Vol.1. New Delhi: Orient Blackswan.

The evolution of retroflex phonotactics in South Asia

Paul Arsenault
Tyndale University College
parsenault@tyndale.ca

1 Introduction

Retroflexion is a well-known areal feature of South Asia. Most South Asian languages, regardless of their genetic affiliation, contrast retroflex consonants with their non-retroflex dental and/or alveolar counterparts (Emeneau 1956; Ramanujan and Masica 1969; Bhat 1973). However, retroflex consonants often exhibit a limited phonotactic distribution relative to their non-retroflex counterparts, a fact that requires explanation. This paper examines two contradictory phonotactic restrictions on retroflex consonants in South Asia and argues that they are best explained if phonotactic restrictions on retroflexion are a direct result of the evolution of retroflexion in a language (or language family). This explanation differs from previous accounts, which posit (possibly universal) synchronic markedness constraints on retroflexion, grounded in speech perception.

2 Typology of retroflex phonotactics

In South Asia, phonotactic restrictions on retroflex consonants are of two basic types. For convenience we can label them the Dravidian and Tibeto-Burman (TB) types. These are summarized in Table 1, where {C_1, C_3} represent pre-vocalic positions and {C_2, C_4} represent post-vocalic positions.[1]

	C_1	V	C_2	C_3	V	C_4
Dravidian type	–		-ʈ	(-ʈ)		-ʈ
TB type	ʈ-		–	ʈ-		–

Table 1. Two retroflex phonotactic patterns

Dravidian languages tend to avoid retroflex (and other apical) consonants in strictly pre-vocalic environments {C_1, C_3}, and favour them in strictly post-vocalic environments {C_2, C_4}. An important caveat is that they typically allow retroflex consonants in C_3 position if C_2 is also retroflex (i.e., homorganic clusters).[2] As a result, the restriction often shows up as a prohibition on word-initial retroflexes (and other apicals). This pattern is

[1] Intervocalic consonants, which are both pre- and post-vocalic, are typically unrestricted for place of articulation. Hence, they are not included in the discussion.

[2] This is typically the result of progressive assimilation, diachronically if not synchronically (e.g., ʈt > ʈʈ, ɳt > ɳʈ).

ascribed to Proto-Dravidian and preserved in most South Dravidian languages. For example, the distribution of coronal consonants in the Kanniyakumari dialect of Tamil is shown in Table 2. Notice that apical alveolar and retroflex consonants are avoided word-initially (C_1) but occur freely in non-initial positions.

	C_1	C_2C_3	C_2C_3	C_4
/t̪/ [n̪]	t̪-, [n̪-]	-t̪t̪-, [-n̪n̪-]	-[n̪]t̪-	—
/t, n, l/	—	-tt-, -nn-, -ll-	-nt-	-n, -l
/ʈ, ɳ, ɭ/	—	-ʈʈ-, -ɳɳ-, -ɭɭ-	-ɳʈ-	-ɳ, -ɭ
/c, ɲ/	c-, ɲ-	-cc-, -ɲɲ-	-ɲc-	—

Table 2. Distribution of coronals in Tamil (Kanniyakumari dialect, Christdas 1988)

The Dravidian pattern is the most common cross-linguistically. Some version of it also occurs in many Indo-Aryan and Australian languages, among others.

Tibeto-Burman languages tend to exhibit a very different pattern. These languages favour retroflex consonants in pre-vocalic environments {C_1, C_3} and avoid them in post-vocalic environments {C_2, C_4} (i.e., syllable codas). For example, the distribution of coronal consonants in Lhomi is shown in Table 3. Notice that retroflexes are favoured in onsets and avoided in codas.

	C_1	C_2	C_3	C_4
/t̪, t̪ʰ/	t̪-, t̪ʰ-	-t̪C-	-Ct̪-	-t̪
/ʈ, ʈʰ/	ʈ-, ʈʰ-	—	-Cʈ-	—

Table 3. Distribution of coronals in Lhomi (Vesalainen and Vesalainen 1976)

The TB pattern is less frequent than the Dravidian pattern cross-linguistically, but is well attested within the TB family.

The Dravidian and TB patterns are contradictory; where retroflexes are avoided in one, they are preferred in the other, and *vice versa*. Under these conditions it is impossible to identify any environment as universally marked or unmarked for retroflexion. However, both patterns can be explained naturally when the evolution of retroflexion within each family is considered.

3 Evolution of retroflex phonotactics

Cross-linguistically, the most common diachronic source of retroflexion is the class of liquids, most notably rhotics (*r*-sounds) but also laterals (*l*-sounds) (Bhat 1973; Hamann 2003, 2005). Liquids are prone to a degree of phonetic retroflexion, especially when they are realized as approximants (e.g., English [ɹ]). They can induce retroflexion in adjacent consonants through assimilation or coalescence. More often than not, retroflexion evolves in a language via progressive assimilation from a liquid to a following consonant (e.g., rt > rʈ > ʈ). However, it can also evolve via regressive assimilation from a liquid to a preceding consonant (e.g., tr > ʈr > ʈ).

The central hypothesis of this paper is that different phonotactic restrictions on retroflexion are a direct result of different evolutionary paths in the development of retroflexion. Specifically, the Dravidian pattern, which avoids initial retroflexion, results directly from retroflexion via progressive assimilation, while the TB pattern, which avoids retroflexion in codas, results from retroflexion via regressive assimilation.

In Proto-Dravidian (PDr), retroflexion first emerged in the class of liquids (**l > PDr *l, *ɭ, *ɻ) and spread from liquids to following nasals and stops via progressive assimilation across morpheme boundaries (Zvelebil 1970; Tikkanen 1999; Levitt 2010). These developments are summarized in (1).

(1) Evolution of coronal contrasts in PDr

a. Starting point	t̪-	-(t̪)t̪-	-lt̪-	-ɭt̪-, -ɻt̪-
b. Progress. assim.	–	–	-lt-	-ɭʈ-, -ɻʈ-
c. Loss of liquid	t̪-	-(t̪)t̪-	-(t)t-	-(ʈ)ʈ-

Ultimately, the loss of the conditioning liquid in (1c) yielded a three-way contrast between dental, apical alveolar and retroflex stops, with the original dental series in both initial and non-initial positions and the new alveolar and retroflex series limited to non-initial environments. Similarly, retroflexion has developed via progressive assimilation from liquids and back vowels in Indo-Aryan (Misra 1967; Bhat 1973; Hamp 1996; Tikkanen 1999) and Australian (Dixon 2002), producing comparable phonotactic patterns in those language families.

Tibeto-Burman languages have developed retroflexion primarily via regressive assimilation in Cr- and Cl- onset clusters (Bhat 1973; Matisoff 2003). The evolution of retroflexion in Tibeto-Burman is summarized in (2).

(2) Evolution of coronal contrasts in TB

a. Starting point	t̪-	Cr-	-t̪
b. Regress. assim.	–	ʈr-	–
c. Loss of liquid	t̪-	ʈ-	-t̪

In the case of TB, loss of the conditioning liquid in (2c) yielded a two-way contrast between dental and retroflex consonants, but only in pre-vocalic onset positions.

In summary, the Dravidian phonotactic pattern emerges whenever retroflexion evolves through progressive assimilation. -rC and -lC sequences are typically non-initial and post-vocalic because they constitute well-formed syllable codas, but not well-

formed onsets. Therefore, progressive assimilation in these clusters produces retroflexion only in non-initial, post-vocalic environments. The TB pattern emerges whenever retroflexion evolves through regressive assimilation. Cr- and Cl- sequences are typically syllable-initial and pre-vocalic because they constitute well-formed syllable onsets, but not well-formed codas. Therefore, regressive assimilation in these clusters produces retroflexion only in syllable onsets.

4 The role of perception

The proposed evolutionary account differs from previous accounts, based primarily on evidence from Australian and Dravidian languages, which argue for perceptually motivated synchronic constraints on retroflexion (Hamilton 1996; Steriade 2001; Hamann 2003). The perceptual cues critical to retroflex contrasts are most salient in VC transitions, and least salient in CV transitions. Thus, according to these accounts, retroflex contrasts are avoided in strictly CV positions, where they lack salient cues, and favoured in VC positions, where they benefit from robust cues.

The perceptual account provides a plausible explanation of the dominant Dravidian type pattern, but fails to explain the TB pattern. In the TB pattern, retroflex segments are restricted to those environments where their cues are least salient, and prohibited in those environments where their cues are most salient. The existence of the TB pattern suggests that the perceptual account cannot be generalized into any kind of universal markedness constraint or implicational universal, such as Steriade's (2001) law of apical contrast, which states that apical or retroflex contrasts can only occur word-initially if they also occur after vowels. Clearly, the TB pattern violates this prediction.

The distribution of perceptual cues may still play an important role in the evolution of retroflex phonotactics. In particular, it may explain the frequency of the Dravidian type pattern relative to the TB pattern. Retroflex assimilation is expected to be predominantly progressive because progressive assimilation preserves the salient VC transition at the expense of the less salient CV transition (VC_2C_3V > VC_2C_2V), whereas regressive assimilation does just the opposite (VC_2C_3V > VC_3C_3V). Thus, the Dravidian pattern, which evolves via progressive assimilation, is expected to be more frequent than the TB pattern, which evolves via regressive assimilation.

5 Conclusion

South Asian languages provide important insight into the origins of retroflex phonotactics. The evidence from South Asia indicates that different (and even contradictory) phonotactic restrictions on retroflexion can emerge as a direct result of the evolution of retroflexion in a language. Specifically, the Dravidian type pattern, which avoids initial retroflexion, results directly from progressive assimilation in liquid-plus-consonant sequences (e.g., -rt > -rʈ > -ʈ), while the TB type pattern, which avoids retroflexion in codas, results directly from regressive assimilation in consonant-plus-liquid sequences (e.g., tr- > ʈr- > ʈ-).

The evidence from South Asia raises doubts about the existence of any universal perceptually motivated synchronic constraints on retroflexion. However, the salience of retroflex cues in VC transitions may explain the fact that progressive retroflex assimilation is more common than regressive retroflex assimilation, and therefore, that the Dravidian type pattern evolves more frequently than the TB type pattern.

References

Bhat, D. N. S. 1973. "Retroflexion: an areal feature". *Working Papers on Language Universals* 13: 27-67.

Christdas, P. 1988. *The phonology and morphology of Tamil*. Doctoral dissertation, Cornell University.

Dixon, R. M. 2002. *Australian languages: Their nature and development.* Cambridge: Cambridge University Press.

Emeneau, M. B. 1956. "India as a linguistic area". *Language* 32 (1): 3-16.

Hamann, S. 2003. *The phonetics and phonology of retroflexes.* Utrecht: LOT.

Hamann, S. 2005. "The diachronic emergence of retroflex segments in three languages". *Link* 15 (1): 29-48.

Hamilton, P. 1996. *Phonetic constraints and markedness in the phonotactics of Australian aboriginal languages.* Doctoral dissertation, University of Toronto.

Hamp, E. P. 1996. "On the Indo-European origins of the retroflexes in Sanskrit". *Journal of the American Oriental Society* 116 (4): 719–723.

Levitt, S. H. 2010. "Fortunatov's law and Dravidian". *International Journal of Dravidian Linguistics* 39 (1): 19–90.

Matisoff, J. A. 2003. *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction.* Berkeley: University of California Press.

Misra, B. G. 1967. *Historical phonology of Modern Standard Hindi: Proto-Indo-European to the present.* Doctoral dissertation, Cornell University.

Ramanujan, A. K., and Masica, Colin. 1969. "Toward a phonological typology of the Indian linguistic area". In T. A. Sebeok (ed.), *Current trends in linguistics, volume 5: Linguistics in South Asia*, 543-577. Paris:

Mouton.

Steriade, D. 2001. "Directional asymmetries in place assimilation: A perceptual account". In E. Hume and K. Johnson (eds.), *The role of speech perception in phonology*, 219-250. New York: Academic Press.

Tikkanen, B. 1999. "Archaeological-linguistic correlations in the formation of retroflex typologies and correlating areal features in South Asia". In R. Blench and M. Spriggs (eds.), *Archaeology and language IV: Language change and cultural transformation*, 138-148. London & New York: Routledge.

Vesalainen, O., and Vesalainen, M. 1976. *Lhomi phonemic summary*. Kathmandu: Summer Institute of Linguistics and Institute of Nepal and Asian Studies.

Zvelebil, K. 1970. *Comparative Dravidian phonology*. The Hague: Mouton.

Compound verbs in Hindi: evidence gained through experimental paradigms

Samiksha Bajpai
National Institute of Advanced Studies, Bangalore

Kiran Kishore
University of Hyderabad

Gautam Sengupta
University of Hyderabad

Rajesh Kasturirangan
National Institute of Advanced Studies, Bangalore

The aim of this paper is to get an insight into how compound verbs in Hindi are accessed and stored in the mental lexicon. We designed two lexical decision tasks to test the hypotheses that compound verbs are cognitively more complex than simple verbs, although they are stored and accessed as a single construction. Verb sequences on the other hand, are cognitively even more complex than compound verbs. These hypotheses were arrived at considering the frequency of their usage in Hindi, which ranges between 15-20% of the total text frequency, as also the theoretical literature in the field. This is the threshold frequency at which any construction ceases to be merely a source of lexical enrichment and takes on a grammatical role instead (Hook, 2001).

Compound Verbs (CVs) are constructions comprising a sequence of two verbs in which the second verb is either delexicalised or its semantic domain is considerably altered. While the semantic centre of gravity varyingly shifts towards the first verb (V1), the major responsibility of bearing the grammatical weight is borne by the second verb (V2). For example, in Hindi, the verb *bol* is 'to speak' and the verb *uTh* is 'to rise'. The compound verb *bol uThA* 'speak rise-sing-pst-m' expresses a manner of speaking, namely to speak out suddenly and without according much thought to what is spoken. Interestingly, while there are few restrictions on the choice of the first verb, the second verb belongs to a set of closed-class elements.

Verb sequences on the other hand, are made up of a sequence of two verbs in which each verb contributes equally to the emergent construction, and they represent two contiguous actions. Thus compound verbs are different from verb sequences.

As noted by Libben (2006), compounds are structures located at the crossroads between morphology and syntax, as they reflect both the properties of linguistic representation in the mind and grammatical processing. The literature on

compound word processing is broadly divided into the decompositional, non-decompositional and the dual access theories, according to which both types of lexical representations (i.e., constituents and whole-words) may exist simultaneously when processing complex and compound words. At the same time, different factors, such as semantic transparency, cumulative frequency, productivity of the affix, and surface frequency determine which lexical representation is activated (Shabani, 2012).

In Cognitive Linguistics, the notion of lexicalization is strongly tied to the so-called 'degree of entrenchment' of a unit. The term 'entrenchment' is aimed to convey the idea that a lot of what speakers say is based on pre-packaged units that are 'entrenched' in our memory so deeply that their activation is practically automatic (Schmid, 2007). If units are entrenched, they have achieved the status of conventional items (Langacker, 1990) and their activation becomes automated to the extent that they have been used before. Therefore, constructions fall "along a continuous scale of entrenchment in cognitive organization" (Langacker, 1987).

In the first experiment, we hypothesised that since CVs are mapped on to a single event as against serial verbs which are mapped onto more than one event, more complex being mapped to more than one event, the time required to process CVs should be less than serial verbs though more than simple verbs. Our experimental findings support this hypothesis and our findings indicate that increased complexity in the representation of verbs leads to increased processing times.

In the second experiment, we ran a lexical decision task using the masked priming technique on CVs in Hindi. We decided on masked priming to ensure that priming is done at the subliminal level (the prime was displayed only for 50 ms) and is therefore free of orthographic, phonological and morphological effects. The priming is therefore semantic. The experiment was designed using 30 different CVs as masked primes and the targets were the respective V1 and V2. The motivation for this experimental design is that if the CV is accessed in terms of its constituents, then post semantic exposure of the CVs, access to V1 and V2 should be facilitated, since the priming is semantic, and the semantics of the CV is essentially the semantics of V1. V2 of a CV, however being delexicalised, there must be no facilitation in the reaction time of the lexical decision task. As a control, we had random verbs which were semantically completely unrelated to the CVs, also prime the same V1 and V2 of the CV.

We used a forward mask: ##### (equal to the length of the prime) centred on the screen for 500 ms after which the prime was displayed for 50 ms. The experiment was self-paced. V1,V2 or the pseudowords were displayed till key-press or for 2 seconds. Subjects asked to do lexical decision task. If the response was incorrect or the subject was over a more than 1 sec, a short beep given.

We found that while there is a significant facilitation in the case of a CV priming V1, there is no significant facilitation or inhibition in the reaction time of a lexical decision reaction task in the case of V2. This suggests that CVs in Hindi are stored as a single cognitive unit or a construction and are not accessed in terms of their constituents.

Thus both experiments seem to point towards the hypothesis that the compound verb in Hindi is stored and accessed as a single cognitive unit, without any decomposition. In this case, Libben's (2006), non-decompositional model holds.

References

Hook, P. 2001. "*Where do compound verbs come from? (And where are they going?)*", The Yearbook of South Asian Languages and Linguistics, pp. 102-130.

Langacker, R.W. 1987. *Foundations of Cognitive Grammar.* Vol. 1: Theoretical Prequisities. Stanford: Stanford University Press.

Langacker, R.W. 1990. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin; New York: de Gruyter.

Libben, G. & Jarema, G. (Eds.) (2006). *The Representation and Processing of Compound Words.* Oxford: Oxford University Press

Schmid, H-J. 2007. *Entrenchment, salience and basic levels.* In D. Geeraerts & H. Cuyckens (eds.), Handbook of Cognitive Linguistics, 169-203. Oxford: Oxford University Press.

Shabani-Jadidi Pouneh. 2012. *Processing compound verbs in Persian.* Unpublished dissertation. University of Ottawa

Possessive agreement markers in Assamese

Gitanjali Bez
Gauhati University
bezgitanjali@gmail.com

This paper is an attempt to discuss the possessive agreement markers in Assamese, a language spoken in the eastern part of India, by a majority of people living in the state of Assam. The discussion is based on a typological survey of possessive agreement markers.

The possessive agreement markers are found to be attached to a kind of possessed nouns in terms of the category of person of the possessor. These markers appear as suffixes and distinguish all the three persons in Assamese. The first person possessor does not require any marker to be added to the possessed noun while the second and third person possessors require different markers. The second person possessor triggers two separate markers on the basis of the distinction of honorificity. *-r* is triggered by the second person inferior possessor while *-ra* is controlled by the second person familiar possessor. The marker *-k* is found to agree with the third person possessor irrespective of the distinction of honorificity. The examples of these markers are given in (1)-(4).

(1) *mʊ-r* *ma-Ø.*
1SG-GEN mother
'My mother.'

(2) *tʊ-r* *ma-**r**.*
2SG.INF-GEN mother-2INF
'Your mother.'

(3) *tʊma-r* *ma-**ra**.*
2SG.FAM-GEN mother-2FAM
'Your mother.'

(4) *ta-r/tɛõ-r* *ma-**k**.*
3SG.INF-GEN/3SG.FAM mother-3
'His mother.'

Cross-linguistically, it is not very uncommon to find person agreement on possessed nouns (Siewierska 2004). But languages vary with respect to their formal coding of person agreement. In some languages, the person agreement markers on possessed nouns are very identical with personal pronominal forms while in others, they are different from personal pronominal forms. Assamese is a language where the possessed noun markers are found to be different from personal pronominal forms as shown in (1-4).

The existence of person agreement on possessed nouns is an unusual phenomenon in Assamese in that it is neither common in NIA languages nor in South Asian languages (Paudyal 2008). However, Darai, an NIA language spoken in Nepal shares a close resemblance with Assamese in both structure and form in this regard as illustrated in (5) and (6).

(5) *te-rʌ* *bʰai-r.*
2SG-POSS brother-2SG
'Your brother.'

(6) *ik-rʌ* *bʰai-k.*
3SG-POSS brother-3SG
'His brother.' (Paudyal ibid)

If the two sets of examples provided from Assamese (1-4) and Darai (5 and 6) are compared, it can be seen that both sets are identical in structure and form. The possessed nouns are marked by person suffixes and preceded by a possessor in the genitive (possessive in Darai) in both languages. Further, the genitive marker and the person agreement show a close affinity in both languages.

The person agreement on possessed nouns is assumed to have evolved from pronominal forms through various historical stages in Darai (Paudyal ibid). But the data found from Assamese show that these suffixes are derived from genitive markers rather than pronominal forms. The reason of this line of thinking is that the markers do not resemble any pronominal form in Assamese. Rather, these are the same as the genitive markers found in some NIA languages, including Assamese (Masica 1991; Dasgupta 2003; Ray 2003; Yadav 2003).

The distribution of different forms of genitive such as *-rɔ* with the possessor, *-r*, *-ra* and *-k* with the second person inferior, familiar and the third person seems to be related to the historical development of the genitive in Assamese. The status of *-r* and *-k* as genitive markers has been discussed by Chatterji (1926). He refers to the word *kara* that was used as a genitive in Magadhi prakrit and Apabhramsa as a source of *-r* and *-k*. According to him, the use of *kara* was originally restricted to pronouns in Magadhi Prakrit and then was extended to nouns. Later, *kara* was reduced to <*-ara* > *-ra*, *-r*> and the reduced forms appeared as genitive affixes with the first and second personal pronouns in languages like Assamese, Bengali, Oriya, Maithili and in other Bihari speeches. The unreduced form *kara* survives as the genitive in Oriya with the plural affix <*-n-*>. It also occurs in the forms <*-kar*, *-karā*, *-kɸrā*> in languages like Maithili, Magahi, Bhojpuriya as the genitive and dative of the pronouns other than the first and second persons. *-kɔrɔ* is still found in Oriya with honorific plural suffix *-ŋ* (Dasgupta 2003). The

occurrence of *-r* and *-k* genitives as agreement markers of the second and third person possessors is attested in the early document of Assamese attributed to 14th century. From this, it is reasonable to assume that their distribution in the second and third person possessed nouns may have developed at the time when NIA languages started to emerge. It is possible that the original system of genitive marking on pronominal forms in terms of person may have shifted further from the pronominal forms to possessed nouns in Assamese. Peterson's (2010) study provides another evidence of the use of *-k* genitive for the third person. This marker occurs with a slight variation in a number of Munda languages along with Sadri from NIA stock spoken in Jharkhand. In both Munda and Sadri languages, the marker that is used to mark some predicative categories in the third person is identical with the genitive marker. The third person marker of this kind is assumed to have evolved from the genitive in those languages.

Typologically, Assamese and Darai form a group with regard to the person coding on possessed items from the perspective of both the structure and form. But the semantic distinction of possessed items sets both these languages apart from each other. The distinction of alienable and inalienable possession is not maintained in Darai. The agreement markers are found to occur with both alienable and inalienable nouns in Darai while in Assamese the markers are triggered only by some kind of inalienable possessed nouns, i.e. kin terms.

The occurrence of person agreement on possessed nouns in Assamese and Darai may be a result of language contact since it is not a common feature of NIA languages as stated above. The major language family which has contact with both Assamese and Darai is the Tibeto-Burman family. However, the person agreement markers found on possessed nouns in TB languages differ from Assamese and Darai in terms of both structure and form. They are used in TB languages as prefixes and derived from pronominal forms (Paudyal 2008; Jacquesson 2008; Chelliah 1997). Thus this phenomenon cannot be a TB influence. Rather, it may be an influence of Munda languages from Austro-Asiatic family. In Munda languages, the agreement markers are attached as suffixes to inalienable nouns as illustrated in (7) and (8). Both examples are cited from Peterson (2010).

Kharia, a South Munda language spoken in Jharkhand

(7) *aba=ɲ*
father=1SG
'My father.

Santali, a North Munda language spoken in Jharkhand

(8) *hɔpɔn=me*
son=2SG
'Your son.'

In both examples, the agreement markers are attached to inalienable nouns, but not with alienable nouns as demonstrated below.

Kharia

(9) *iɲ=aʔ* *khoɽi*
1SG=GEN village
'My village'

Santali

(10) *am=ak'* *oɽak'*
2SG=GEN house
'Your house.' (Peterson ibid)

Examples from Munda languages cited above are similar with Assamese in two respects: (i) the person agreement appears as suffixes in both languages, and (ii) the alienable-inalienable distinction is maintained in both languages. The Munda languages may have influenced Assamese in these two respects. At the same time, both languages show other differences. For instance, the inalienable noun is not preceded by a possessor in the genitive in Munda languages while in Assamese, it is. Secondly, the possessive agreement markers seem to be identical with personal pronominal forms (*ɲ~ iɲ* and *me~am*) in Munda languages while in Assamese, they are identical with genitive markers.

From the data stated above, it can be concluded that Assamese shows a unique pattern with regard to the person agreement on possessed nouns. It neither displays complete resemblance with Darai nor with Munda languages, but seems to have internally developed a new pattern which is unusual among the languages of this region.

Acknowledgements

I would like to express my gratitude to Prof. Jyotiprakash Tamuli for encouraging me to prepare this abstract. I would also like to offer my sincere thanks to Prof. Scott DeLancey for his insightful comments and suggestions on the topic. The responsibility for all errors and inaccuracies in this work, are, of course, mine.

Abbreviations

1 first person
2 second person
3 third person

FAM familiar
GEN genitive
INF inferior
POSS possessive

References

Chatterji, S. K. 1926. The Origin and Development of the Bengali Languages. New Delhi, Rupa Co, India.

Chelliah, S. L. 1997. A grammar of Meithei. Berlin, Walter de Gruyter.

Dasgupta, P. 2003. "Bangla." In G. Cardona and D. Jain, Eds. The Indo-Aryan Languages. London/New York, Routledge:351-390.

Jacquesson, F. 2008. A Kokborok Grammar. Tripura. Kokborok Tei Hukumu Mission.

Masica, C. P. 1991. The Indo-Aryan Languages. Cambridge/Port Chester/Melbourne/Sydney, Cambridge University Press.

Paudyal, N. 2008. "Agreement patterns in Darai: typological study." Nepalese Linguistics. Vol. 23, 186-207.

Peterson, J. 2010. "Language contact in Jharkhand. Linguistic convergence between Munda and Indo-Aryan in eastern-central India." Himalayan Linguistics 9/2:56-86. http://www.linguistics.ucsb.edu/HimalayanLinguistics/articles/2010/PDF/HLJ0902B.pdf

Ray, T. S. 2003. "Oriya." In G. Cardona and D. Jain, Eds. The Indo-Aryan Languages. London/New York, Routledge: 444-476.

Yadav, R. 2003. "Maithili." In G. Cardona and D. Jain, Eds. The Indo-Aryan Languages. London/New York, Routledge: 477-497.

A quantitative study of Bodo agglutination

Krishna Boro
University of Oregon
krishna@uoregon.edu

Prafulla Basumatary
Gauhati University
pbasumatari@gmail.com

1 Introduction

Bodo is a Tibeto-Burman language spoken mainly in the state of Assam in India. The verb morphology is highly agglutinating. More than hundred **stem-formatives** and around two dozen **inflectional morphemes** form part of the verb morphology. These stem-formatives and the inflectional morphemes are concatenated to create a large number of unique polymorphemic words and word-forms. For instance, the verb lexeme *nai* 'see' has 602 unique word-forms in our database. Word-forms containing 1 to 3 affixes are very common in our database. The longest verb form we have encountered in our database so far contains five affixes.

We conduct a corpus-based quantitative study of the agglutinating verb morphology via a frequency and co-occurrence analysis. We concentrate on the **stem-formatives** since they are quite large in number and theoretically they can be combined into extremely long sequences. The frequency analysis consists of counting the frequency of each of the stem-formatives in our database using the concordance feature of FLEx.[3] The co-occurrence analysis investigates frequent co-occurrence of specific stem-formatives suffixes. In and of itself, the product of this study contributes towards a quantitative description of the verb morphology. However, a greater implication of this study is in its role in reflecting an intuition about the ordering of the affixes in a verb form. Intuitively, it seems that it is the nature of the semantic content of the affixes that determines the order – the more specific the meaning, the closer the affix to the root. However, it is difficult to quantify this intuition. We believe that **frequency distribution** and **co-occurrence analysis** will better reflect this intuition and the ordering between stem-formatives themselves.

2 Verb morphology

The verb allows concatenation of multiple morphemes in a string of suffixes and prefixes, which we can easily separate and assign some kind of semantic characterization to, some of which are highly specific while others are quite abstract in

[3] For details on FLEx, please visit http://fieldworks.sil.org/flex/

nature. The following example illustrates a verb form in our database.

1. *mɯider-a* *bi-sur-kʰou* *sigi-nanɯi*
   elephant-NOM 3-PL-ACC frighten-NF
   hor-ou ***hú-kʰár-hór-pʰin-gar-dɯŋ***.
   night-LOC drive-MOT-DIST-REPT-REGRET-RLS
   'The elephant scared and drove them away again unfortunately.' [BWC 108:2.23]

In example (1) the finite verb consists of the verb root ***hú*** 'drive away' and five other morphemes. Four of these are transparently verbal in origin, though some of them have acquired quite abstract meaning. The morphemes ***kʰár, hór, pʰin,*** and ***gar*** lexically mean 'to run', 'to send something to somebody', 'to reply', and 'to discard' respectively. However, as suffixes they indicate 'a motion, which does not have to involve running with legs', 'a distal motion instead of a proximal', 'repetition', and 'a sense of regret' respectively. These morphemes may be regarded as **derivational** or **stem formatives**, in that they do not create 'fully functional' verb forms. The morpheme ***dɯŋ*** is inflectional, i.e. it creates 'fully functional' verb forms, and indicates a realis mood.

The bulk of stem-formatives are what are called **Adverbial suffixes** (see Burling 2004) like *-kʰár,-hór,-pʰin,-gar* in (1). The number of Adverbial suffixes is well above hundred, may even be two hundred. There are also a couple of causative prefixes. Since, Adverbial suffixes are what characterizes verb morphology as agglutinating, our study mainly concentrates on them.

The Adverbial affixes are an interesting set of morphemes with diverse historical origins. A subset of these morphemes is serial verbs, which are grammaticalized forms a set of verbs which still exist in the language as lexical verbs (Boro 2012). We have seen examples of serial verbs in example (1). There are more than a dozen of these serial verbs. Some frequent ones are given in Table 1.

Serial verbs	**Grammaticalized meaning**	**Lexical meaning**
zɯb	'Exhaustive'	'end'
kʰaŋ	'Completive'	'put a pot down after cooking'
kʰaŋ	'Upward movement'	'raising a baby'
hɯi	'Distal motion with event at the end of the motion'	'go and give'
laŋ	'Distal motion with event somewhere along the motion'	'take away'

Table 1: Some of the serial verbs in Bodo

A subset of the Adverbial suffixes are sound symbolic, morphemes that are somehow associated directly with the content meaning in the speakers' mind.

2. *bi-sur* *hagra-boŋgra* *boŋpʰaŋ*
   3-PL jungle-RED tree
   sa-mani *bu-**zrao**-bai-ja-swi.*
   top-EVEN beat-SOUND.SYM-CASUAL-REP-CS
   'They beat around the bush, and even tree tops.
   [BWC 19:2.58]

In (2), the Adverbial suffix *zrao* imitates the sound of hitting a hard surface with something leafy, such as a branch, broom, etc. For most part of the Adverbial suffixes, the origin is unknown.

3 Method

Our database consists of both spoken and written data. The written data consists of 975 texts containing 1.1 million words (tokens). The spoken data consist of 101 texts containing around 100,000 words (tokens). The database contains texts from various genres – such as conversation, narrative, personal accounts, formal speech, procedural, etc.

The frequency analysis consists of counting the frequency of each of the stem-formatives in our database using the concordance feature of FLEx. The co-occurrence analysis mainly follows Durrant's (2013) method of **study of formulaicity**. The co-occurrence analysis investigates frequent co-occurrence of specific Adverbial suffixes. We choose inflected forms of **twenty verbs** with a wide range of different frequencies as our database for co-occurrence analysis. Within this database, we select those stem-formatives which have a **frequency of 100 or higher** for co-occurrence analysis.

4 The implications of the study

Besides providing a frequency distribution and co-occurrence analysis of the verb morphology, this study, among other things, may shed light on **the problem of ordering** of the stem-formatives (mostly Adverbial suffixes) with respect to the verb root and with respect to each other. First, the frequency analysis can provide an insight into the problem of ordering Adverbial suffixes with respect to the root. Intuitively, it seems that Adverbial affixes which have more concrete content meaning tend to appear close to the verb root, and Adverbial affixes which have more abstract meaning tend to appear further away from the root. An informal investigation of frequency distribution of twelve Adverbial suffixes in Table 2 shows that adverbial suffixes with concrete meaning have low frequency, while Adverbial suffixes with abstract meaning have

high frequency. Thus, in terms of frequency it appears that low frequency Adverbial suffixes appear closer to the root than the high frequency Adverbial suffixes. Thus, we may find a strong correlation between the position of the Adverbial suffixes within a verb form and their frequency.

Adverbial suffix	**Meaning (basic)**	**Frequency**
-sling	cut diagonally, on a slant	21
-kʰleng	to make something bend	31
-kʰlab	detaching a small part of something	66
-glaŋ	do something leasurely	71
-pʰer	to have an experience in/of something	73
-kʰao	to split in half	95
-gluŋ	do something chaotically (such as beating unscrupulously)	100
-kʰlai	to make something fall down	134
-pʰla	to pretend to do something	230
-pʰlaŋ	to do something by mistake	249
-pʰui	proximal motion	300
-pʰin	repetition	1070

Table 2: Frequency distribution of some Adverbial suffixes

Second, the co-occurrence analysis will provide an insight into the problem of ordering between Adverbial suffixes. The order between the adverbial suffixes is by no means fixed, as illustrated by examples (3) and (4), where we find both the order *hor-pʰin* 'distal-repetitive' and the order *pʰin-hor* 'repetitive-distal'.

3. *gosai-a* *aŋ-kʰou* *nai-**hor-pʰin**-duŋ.*
   god-NOM 1SG-ACC look-DIST-REPT-RLS
   'God looked back at me.' [108:2.292]
4. *bi-sur* *nai-**pʰin-hor**-lai-duŋ.*
   3-PL look-REPT-DIST-RECP-RLS
   'They looked back at each other.' [493:2.142]

It is highly likely that one of the possible orders is more preferred, as is the case with *hor-pʰin* with 25 counts and *pʰin-hor* with only 7 counts in our database. This indicates that one of the orders is more formulaic than the other. The co-occurrence analysis will provide a more tangible answer to the ordering problem between the stem-formatives.

Abbreviations

1	first person
2	second person
3	third person
ACC	accusative
AFF	affirmative
CS	change of state
DIST	distal
GEN	genitive
LOC	locative
MOT	motion
NF	non-finite
NOM	nominative
PL	plural
RECP	reciprocal
RED	reduplication
REP	reportative
REPT	repetitive
RLS	realis
SG	singular
SOUND.SYM	sound symbolic

References

Boro, K. 2012. "Serialized Verbs in Boro". In G. Hyslop, S. Morey and M. Post (eds.). *North East Indian Linguistics 4.* New Delhi: Cambridge University Press

Burling, R. 2004. *The Language of the Modhupur Mandi (Garo).* Vol. I. New Delhi: Bibliophile South Asia.

Durrant, P. 2013. "Formulaicity in an agglutinating language: the case of Turkish." *Corpus Linguistics and Linguistic Theory.* Vol. 9 (1):1-38.

Vector verbs in Malabar Indo-Portuguese

Hugo C. Cardoso
Universidade de Lisboa
hcardoso@clul.ul.pt

1 Introduction

South Asian languages are rich in complex-predicate phenomena and, in this domain, vector verbs have received much attention in the literature. Vector verbs are, in fact, often seen to constitute one of the most relevant characteristics of the South Asian *Sprachbund* (see e.g. Hook 1974; Masica 1975; Abbi & Gopalakrishnan 1991; Abbi 2001; Butt 2003).

But while vector verbs are well-established in descriptions of the long-standing languages of the region, their presence in more recent newcomers to the region is less well-known (but see Nordhoff 2012 for Sri Lanka Malay). One of these is Indo-Portuguese, which refers to a string of Portuguese-lexified creoles that formed in coastal South Asia from 1498 onwards and currently subsist in Diu, Daman, Korlai and Kerala (India), and Trincomalee and Batticaloa (Sri Lanka). There is a significant degree of differentiation between the several varieties of Indo-Portuguese, but vector verbs are robust at least in those of the former Malabar [i.e. Kerala], which have Malayalam as their substrate/adstrate.[4] Despite its severe endangerment, Malabar Indo-Portuguese is still spoken in Cannanore and was until recently spoken in Vypeen (Cochin). It is currently in the process of being documented and described.

2 Vector verbs

'Vector verbs' have several alternative names in the literature, including 'explicator compound verbs' (e.g. Abbi & Gopalakrishnan 1991; Abbi 2001) and 'light verbs' (Butt 2033; Butt & Lahiri 2013) – and are sometimes also interpreted as 'auxiliaries'. They refer to a limited set of verbs that participate in complex monoclausal (V + V) constructions, in which they typically occur adjacent to the main verb and in some way modulate the predication expressed by it. In these contexts, by definition, vector verbs do not carry the exact semantics that they would when used as main predicates, instead conveying meanings of a more grammatical or pragmatic nature, including such notions as benefaction, volitionality, completion, etc. Two examples from Malayalam are given in (1):

(1) a. Malayalam [Asher & Kumari 1997:347]
aaʃaari meeʃakaɭ ***uŋʈakki taɭɭi.***
carpenter table-PL make-PP push-PST
'The carpenter made the tables (intensively).'

b. Malayalam [Asher & Kumari 1997:348]
ɲaan avaɳɳə raŋʈə pustakam
1s 3s.m-DAT two book
ayaccukoʈ uttu.
send-give-PST
'I sent him two books (for his benefit).'

In (1), only the first verb in the sequence indicates the actual event, while the second one modulates its meaning: in (1a), *taɭ ɭ uka* 'to push' carries a meaning of intensiveness or swiftness; in (1b), *koʈ ukkuka*[5] 'to give' is used to add a notion of benefaction.

Constructions of this type are frequent and robust in Dravidian languages, such as Malayalam or Tamil. In fact, the wide distribution of vector verbs in South Asia, which cuts across language family boundaries, has been given a contact explanation according to which they would have developed in Dravidian languages and then spread to Indo-Aryan languages (see e.g. Steever 1988) – although the exact source and directionality of spread remain disputed.

3 Vector verbs in Malabar Indo-Portuguese

This study of Malabar Indo-Portuguese is based on a corpus collected in Cannanore and Vypeen (Cochin) between 2006 and 2011. In this corpus, several instances of multi-verb constructions occur. In many of these, the secondary verb occurs at the end of the clause, adjacent to the main verb, and contributes exactly the type of modulation of the predication that we recognise from the study of vector verbs in other South Asian languages. Some examples follow:

(2) a. [Vypeen; free-flowing speech, field data]
akə sister paymi hotel-dẽtrə ũga
DEM sister 1s.ACC hotel-LOC one
sirvis ***ja oʌ a da.***
job PST look give
'That sister found a job at a hotel (for me).'

[4] Ian Smith's published corpus of Sri Lanka Portuguese (Smith 1977), collected in Batticaloa, also records the presence of some vector verbs.

[5] The authors fuse the two verbal elements in their transliteration, but clarify in the text that *koʈ ukkuka* is a lexical verb.

b. [Vypeen; elicitation, field data]
boz Goa ***jə*** ***foy*** ***tẽ.***
2s Goa PST go.PST be
'Have you ever been to Goa?'

c. [Cannanore; free-flowing speech, field data]
prẽdisə-pə jə foy ɔ rzə,
learning-to PST go.PST when
purtəgez ***jə*** ***iskusew*** ***foy.***
portuguese PST forget.PST go.PST
'When we went for learning [i.e. school], (we unfortunately) forgot Portuguese.'

The vector verbs at the end of each of these example sentences, all derived from Portuguese etyma, are formally equivalent to full verbs meaning 'to give' (2a), 'to be' (2b) and 'to go' (2c), but their semantics is clearly somewhat bleached here. Predication consists of a single event expressed by the main verb – which precedes them –, and the vector verbs contribute notions of benefaction, completion and pejoration respectively. In these constructions, the duplication of TAM marking on the main and vector verbs is not necessary but also not impossible, as shown in (3):

(3) [Vypeen; free-flowing speech, field data]
P. P. ***jə*** ***more jə*** ***foy*** *agɔ rə.*
P. P. PST die PST go.PST now
'(Sadly,) P.P. has already died.'

Vector verb constructions do not constitute the only type of multi-verb predicates in Malabar Indo-Portuguese, but they do have some features that set them apart. Word order is relevant to distinguish between vector verbs (occurring to the right of the main verb) and a different type of verb which occurs to the left of the main verb, a position reserved for TAM markers – whether they are verbal in nature (4a) or not (4b):

(4) a. [Cannanore; elicitation, field data]
yo butikɑ ***kerɑ*** ***vay*** *peyxi pu pidi.*
1s shop want go fish to buy
'I must go to the shop to buy fish.'

b. [Vypeen; elicitation, field data]
yo minhə irmẽ-pærtə ***lo*** ***mãda.***
1s 1s.GEN sister-near FUT send
'I will send (it) to my sister.'

In (4b), the future particle *lo*, which is not a verbal element, occurs pre-verbally; this position is replicated in (4a) by the verb *kerɑ* , which can also be used as a main verb meaning 'to want' but here contributes a deontic modal meaning (viz. obligation) to the predication. I interpret these as instances of auxiliary verb modification. In addition to the different position of these auxiliaries and vector verbs relative to the main verb, auxiliary constructions as in (4) admit only one instance of the relevant TAM marking.

For these reasons, it is possible to posit a special category of vector verbs in Malabar Indo-Portuguese. Diachronically, vector verbs in Malabar Indo-Portuguese must constitute a local development drawing on local (Dravidian) linguistic strategies, given that Portuguese does not provide a suitable model to account for this linguistic resource. The history of contact indicates that the time-depth of this process must be less than 5 centuries. As a matter of fact, one cannot be sure of whether vector verbs have been a feature of Malabar Indo-Portuguese from the very moment of its formation or whether they entered the language later, as part of a process of convergence towards Malayalam. The few historical records of Malabar Indo-Portuguese available (Schuchardt 1882, 1889) do not contain any instances of vector verbs, but there is good reason to be suspicious of the validity and comprehensiveness of the data contained therein.

References

Abbi, A. 2001. *A manual of linguistic field work and structures of Indian languages*. Munich: Lincom Europa.

Abbi, A. & D. Gopalakrishnan. 1991. Semantics of explicator compound verbs in South Asian languages. *Language Sciences* 1.2: 161-180.

Asher, R.E., and Kumari, T.C.. 1997. *Malayalam* [Descriptive Grammars Series]. London/New York: Routledge.

Butt, M. 2003. The light verb jungle. In G. Aygen, C. Bowern and C. Quinn (eds.), *Harvard Working Papers in Linguistics* 9: 1-49.

Butt, M. & A. Lahiri. 2013. Diachronic pertinacity of light verbs. *Lingua* 135: 7-29.

Hook, P.E. 1974. *The compound verb in Hindi*. Ann Arbor: Center for South & Southeast Asian Studies, University of Michigan.

Masica, C. 1976. *Defining a Linguistic Area: South Asia*. Chicago: University of Chicago Press.

Nordhoff, S. 2012. Multi-verb constructions in Sri Lanka Malay. *Journal of Pidgin and Creole Languages* 27.2: 303-342.

Schuchardt, H. 1882 Kreolische Studien II. Über das Indoportugiesische von Cochim. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Wien (Philosophisch-historische Klasse)* 102: 799-816.

Schuchardt, H. 1889. Beiträge zur Kenntnis des kreolischen Romanisch: VI. Zum Indoportugiesischen von Mahé und Cannanore. *Zeitschrift für Romanische Philologie* 13: 516-524.

Smith, I.R. 1977. Sri Lanka Creole Portuguese Phonology. PhD diss., Cornell University.

Steever, S.B. 1988. *The serial verb formation in the Dravidian languages*. Delhi: Motilal Banarsidass.

Child language: speech of Hindi-speaking infants

Hem Chandra Pande
Jawaharlal Nehru University
hcpande1@rediffmail.com

1 Introduction

Language acquisition by infants has always been a fascinating topic for researchers. Most of the children do develop passive language capability much earlier than they are able to demonstrate their active language capability (ALC). Passive language capability (PLC) means the ability of the infant to understand and follow the oral instructions given by parents and other elders. We are here concerned with the stage when infants having passed the age of cooing and babbling start forming words with speech sounds which they have already acquired. This stage comes at the age of 2+. At this age they also start forming sentences.

The examples of language development of Hindi-speaking infants being given here have been collected from the speech of children between the age of two and two and a half to four-plus (2-2½ to 4+) years. It has been observed that around the age of three years the speech of infants starts developing very fast, and around the age of four years (4±) their language development is almost complete.

We come to know about the language development of children through the errors made by children in the pronunciation and use of words and through the sentences framed by them. Gradually errors in the speech of the infants get reduced, and we come to know about the progress made by infants in language acquisition. Before demonstrating their ALC, all the children are found to be capable of understanding and following oral language instructions which we have called PLC. Infants start comprehending oral instructions from the age of 1½. But ALC comes much later. The progress made by children manifests itself when they try to express their 'thoughts' in the form of incomplete and sometimes incomprehensible words. The meaning of such incomplete words becomes clear from the context. Such 'words' are, in fact, a complete sentence in the child language. The whole word can be represented by the initial or the final part of the word, as we will see in the examples being given below.

2 Acquisition of sounds

Coming to the actual speech of the Hindi-speaking infants as observed by us, we would like to point out that most of the introductory books of linguistics and some books on child language and psycholinguistics (see, for example: Akmajian et al. 1988, pp. 465-479; Fromkin & Rodman 1986, pp. 325-342; Sakharny 1989, 60-64; Yule 2009, pp. 149-161; Shastri 1973, 289-336; Bhatia 1980, pp. 26-29; Shastri & Sharma 1986, 69-71), while dealing with the child language, do not mention about the phonetic features which have been observed by us in the speech of the infants.

Our observations show that initially the infants learn to pronounce vowel sounds. They first learn to utter the vowels [e], [u], [i] followed by the vowels [a] and [o]. The infants learn to pronounce the consonants a little later because the speech organs - lips, teeth, tongue, throat, nose palate, larynx etc. – which participate in the articulation of the consonants, are not yet fully developed, and their movement is restricted. Out of the various consonants of Hindi the infants take much more time to learn to pronounce correctly aspirated consonants, voiceless as well as voiced, and retroflex consonants. The order of their acquisition is roughly like this:

Voiceless aspirates: k^h (ख), c^h (छ), t^h (थ), p^h (फ)
Voiced aspirates: g^h (घ), j^h (झ), d^h (ध), b^h (भ)
Voiceless retroflex: T (ट), T^h (ठ)
Voiced retroflex: D (ड), D^h (ढ)
Nasal retroflex: N (ण)

3 Phonetic errors

Phonetic errors found in the speech of infants can be classified into six types: 1) Simplification, 2) Metathesis, 3) Assimilation, 4) Reduplication, 5) Lengthening of vowels, and 6) Replacement of final [r] with a short [ə]. Here we are giving examples of the first four types only.

Simplification is observed in the form of: a) reduction of syllables, b) deaspiration, and c) dropping of the second consonant in some consonant clusters.

a) Reduction of syllables. The infants are not able to pronounce all the words correctly. One of their difficulties is concerned with the length of the words consisting of three or four syllables which they reduce to two syllables. For example: *kabUtar* 'pigeon' → *kabU*; *darvAjA* 'door' → *dAjA*.

b) Deaspiration. The infants while uttering words having aspirated consonants replace them with the corresponding unaspirated consonants. For example: *k^holo* 'Open' → *kolo* ([k^h] is replaced by [k]); *d^hUl* 'dust' → *dUl* (d^h] is replaced by [d]).

c) Dropping of the second consonant in some consonant clusters. Pronunciation of consonant clusters requires special efforts. Therefore, children tend to simplify them by dropping the second

consonant in certain clusters. For example: *kyA* 'what' → *ka* ([y] is dropped); *krIm* 'cream' → *kIm* ([r] is dropped).

Metathesis of consonants is very common in the speech of the infants (Malmberg 1963, p. 63). Here are some examples: *godI* 'lap' → *dogI* ([g] and [d] have been interchanged); *TopI* 'cap' → *poTI* ([T] and [p] have been interchanged).

Assimilation is also quite common in child language. Here are some examples: *kap* 'cup' → [*pap*; *biskiT* 'biscuit' → *bikkiT*; *davAI* 'medicine' → *dadAI*.

The speech of the infants is also characterized by the use of reduplicated words and forms, for example: *cAy* 'tea' → *cAy-cAy*. While playing with various objects the children scatter them all over the bed or room, and then they say p^hEle- p^hEle '[I have] scattered' instead of *phElA diyA*. This is also an instance of reduplication.

4 Grammatical errors

When the infants start speaking our attention is drawn to the grammatical errors made by them in word-forms, phrases and sentences. Most of such errors are concerned with grammatical markers which the children tend to drop. This is a common feature found in the speech of the children. The English-speaking children drop the plural marker *-s*, the article *the*, the preposition *of*, the past tense marker *-ed* (*How Babies Think* –Gopnik et al. 2001, p. 118). This may due to the abstractness of the grammatical markers unlike other words which refer to concrete objects, actions, qualities etc. Some examples of the grammatical errors found in the speech of the Hindi-speaking infants are being given below.

The postpositions *kA*, *kI*, *ke* in the meaning 'of', *ko* in the meaning 'to', me^n in the meaning 'in' occur very frequently. The child while uttering phrases which require the use of these postpositions omits them and uses the nouns in their proper order which helps to convey the desired meaning. For example: *dAdA ke* g^har me^n 'in paternal grandfather's house' → *dAdA* g^har; *mAnI ko de do* 'Give to Mani' → *mAnI de do*. In these examples the postpositions *ke*, me^n and *ko* have been omitted.

Small children are not able to conjugate the verbs. As such they use the infinitive from of the verbs. For example: *mammA, jaldI* $D^hu^nD^ho$ 'Mamma, please, search out fast' → *mammA, jaldI* $D^hu^nD^hnA$. Here the imperative form $D^hu^nD^ho$ has been replaced by the infinitive $D^hu^nD^hnA$. It is rather strange that the children learn to use the infinitive which in the speech of the adults does not occur so frequently.

The infants learn to correctly use the future tense much later. Initially they use the infinitive form, for example: *kal* AU^ngA 'I will come tomorrow'→ *kal AnA*. After some time they start using participial form which is also the present perfect form in Hindi. For example: *mammA kab AegI* 'When will Mamma come?' → *mammA kab AI*.

5 Conclusions

Thus, by constantly observing the speech of the infants we can notice the features of the child language, and how they gradually acquire phonetics and grammar. The general trend of development of speech in infants is identical in all languages. But it is not linear in the sense that growth of speech of the infants is not uniform. Grammar is acquired gradually. After acquisition of grammar the children start communicating actively, and they become more and more expressive.

Finally we would like to mention that the infants are able to convey their desires and thoughts with the help of incomplete words and without fully acquiring the language when they have not yet acquired ALC. The language behaviour of infants indicates that thought can emerge at least without ALC. The language that the child has already acquired in his subconscious helps him to frame his thoughts even though he is not able to express them correctly.

References

Akmajian, A., Demers, R.A and Harnish, R.M 1988. *Linguistics: An Introduction to Language and Communication*, The MIT Press.

Bhatia, Sugan 1980. *Balak men Bhasha ka Vikas*, Kendriya Hindi Sansthan, Agra. (In Hindi)

Chukovsky, Kornei 1970. *From Two to Five*, "Detskaya literatura", Moscow. (In Russian)

Fromkin, V. and Rodman, R. 1986. *An Introduction to Language*, HRW International Edition.

Gopnik, Alison, Meltzoff, A. and Kuhl, P. (2001). *How Babies Think*, Phœnix.

Lennberg, B.S. 1980. *Biological Foundations of Language*, Wiley, New York.

Malmberg, B. 1963. *Phonetics*, Dover Publications, New York.

Sakharny, L.V. 1989. *Introduction to Psycholinguistics*, Izdatelstvo Leningradskogo universiteta, Leningrad. (In Russian)

Shastri, Sitaram and Sharma, Vashini 1986. *Manobhasha Vikas*, Kendriya Hindi Sansthan, Agra. (In Hindi)

Shastri, Suryadev 1973. *Manobhashiki*, Bihar Hindi Grantha Akademi, Patna. (In Hindi)

Yule, G. 2009. *The Study of Language*, Cambridge University Press.

On the nature of objects in unergatives

Pritha Chandra **Gurmeet Kaur**
Indian Institute of Technology, Delhi
prithachandra@gmail.com gurmeetk88@gmail.com

Unergative verbs, across split ergative languages, manifest a lot of variation in terms of allowing an ergative subject. In Hindi-Urdu, unergative predicates 'to bark', 'to sneeze' occur optionally with an ergative subject (Mohanan 1994 and Mahajan 2012), while 'to sing' necessarily occurs with an ergative subject. Similarly, in Basque, unergatives 'to dance', 'to cry' may take an ergative subject (Bobaljik 1994; Preminger 2012), but 'to speak' must take it obligatorily. While some scholars attribute this variation to volitionality of the agent (Mohanan, 1994); others explain it via alternating transitivity of the lexical unergative predicate, (Mahajan, 2012).This paper attempts to contribute to the existing literature on variation in unergatives by proposing a plausible motivation for the varied ergative patterns. Employing data on unergatives from Punjabi, we propose that the presence of an object in an unergative construction is not the sole determinant of morphological ergativity. These structural factors in combination with semantic ones can account for variation better. Specifically, we contend that unergative verbs whose object is ontologically differentiated from the event referred to by the verb, take an obligatory ergative subject. This class of verbs is thus closer to transitives, which also take ergative subjects obligatorily. On the other hand, the lack of such differentiation between the object and the action denoted by the verb makes an ergative subject unavailable in other unergative predicates.

We take four unergative verbs in Punjabi viz. 'to laugh', 'to dance', 'to spit' and 'to scratch', which behave differently with regard to the phenomenon of ergativity. While 'to spit' and 'to scratch' (Type I) require an obligatory ergative subject (1); verbs 'to dance' and 'to laugh' (Type II) require it optionally (2). We pick one verb of each type for illustration.

1. *raam-ne/*raam* *thukkeyaa*
   Ram-erg/*Ram.nom spit.perf.m.sg
   'Ram spit.'
2. *ram /ram-ne* *nacceyaa*
   ram-nom/ram-erg dance.perf.m.sg
   'Ram danced.'

To confirm that we are dealing with verbs of the same class, we probe into the nature of the verbs under discussion. Bhatt (2003), Bobaljik (1993) and Richa (2008) in their discussions on the verb classes in Hindi-Urdu and Basque, provide a set of diagnostics to distinguish unaccusative predicates from the unergatives. We employ these tests to demonstrate that 'to laugh', 'to dance', 'to spit' and 'to scratch' in Punjabi are unergatives. This is confirmed by (i) the presence of a cognate object (3-4); (ii) the formation of an inabilitative passive, as in (5-6); (iii) the use of a perfective participle as a reduced relative, albeit with differences(7-8); (iv) passive formation (9-10) and (v) the obligatory genitive marker on the agent with the imperfective participle of an unergative (11-12).

3. *ram-ne* *thukk* *thukkii*
   ram-erg spit.f.sg spit.perf.m.sg
   'Ram spat a spit.'
4. *siitaa-ne/siitaa* *maad^{h}urii-daa* *nacc* *nacceyaa/nacci*
   sita-erg/sita.nom madhuri-gen.m dance.m dance.perf.m.sg/f.sg
   'Sita danced Madhuri's dance.'
5. *raam-tõ saRak te thukkeyaa nayii geyaa*
   ram-inst road on spit neg go.perf.m.sg
   'Ram was not able to spit on the road.'
6. *siitaa-tõ nacceyaa nayii geyaa*
   sita-inst dance neg go.perfm.sg
   'Sita was not able to dance.'
7. *thukkii* *hoyii* *gitak*
   spit.perf.f.sg happen.perf.f.sg seed.f.sg
   'The spit seed.'
8. **nacceyaa* *hoyaa* *aadmii*
   dance.perf.m.sg happen.perf.m.sg man
   'the danced man.'
9. *gitak-nuũ* *thukkeya* *gayaa* *sii*
   seed-acc spit.perf.m.sg go.perf.m.sg be.pst
   'The seed was spat.'
10. *vyaah vicc nacceyaa gayaa sii*
    wedding in dance.perf.m.sg go.perf.m.sg be.pst
    'Dance happened in the wedding (lit. In the wedding, it was danced).'
11. *raam-*(de) saRak te thukkde ii police aa gayii*
    ram-gen road on scratch.prs.part emph police come go.perf.f.sg
    'Police arrived as soon as Ram spat on the road.'
12. *raam- *(de) naccde ii, rolaa pɛ geyaa*
    ram-gen dance.prs.part emph noise fall go.perf.m.sg
    'As soon as Ram danced, there was a lot of noise.'

These findings have been summarized in table I.

	Type I	Type II
Cognate object	✓	✓
Inabilitative passives	✓	✓
Reduced relatives(perf)	✓	X
Passives	✓	✓
Obligatory genitive marker	✓	✓

Table I

As the verbs of both types manifest unergative behavior, the differential subject marking cannot be attributed to these structural criteria. Further, in keeping with the Implicit Object Conjecture, the unergatives of both types are also underlying transitive, ruling out transitivity as the sole determinant of ergativity. Adjectives can be employed in these constructions irrespective of the presence/absence of an overt object. Consider examples (13-14).

13. *mariiz-ne gaŋḍaa ʈhukkeyaa* / *mariiz-ne gaŋḍii thukk ʈhukkii*
patient-erg dirty.m.sg spit.perf.m.sg/patient-erg dirty.f.sg spit.f.sg spit.perf.f.sg
'The patient spat a dirty spit.'
14. *siiʈaa-ne/sitaa pyaaraa (nacc) nacceyaa/nacci*
Sita-erg/sita.nom lovely.m.sg dance.m.sg dance.perf.m.sg/f.sg
'Sita danced a lovely dance.'

The examples presented so far show that all verbs in question are unergatives that are underlying transitive, which leaves the variation in morphological ergativity unexplained.

We propose a semantic explanation for the presence/absence of the ergative subject with these unergatives based on the nature of the object. Specifically, we contend that Type I verbs are like 'true' transitives ('to hit', 'to kill') in the sense that these predicates have internal arguments that signify objects ontologically differentiated from actions depicted by verbs. To elaborate, transitives like 'to hit' select an object that can be identified as separate from the event of hitting. Consider (15) for illustration, where the object 'child' refers to an entity affected by, but distinct from the action of 'hitting'. We propose that 'to spit' and 'to scratch' fall in this category. Thus, the object 'spit' predicated of 'to spit' in (16) is like the object of a true transitive verb. This argument is further exemplified by 'to scratch' in (17), which can only select an externally individuated entity like 'wound' or 'skin'.

15. *tiicar-ne bacce-nũũ maareyaa*
teacher-erg child-acc hit.perf.m.sg
'The teacher hit the child.'
16. *raam-ne thukk thukkii*
ram-erg spit.f.sg spit.perf.m.sg
'Ram spat a spit.'
17. *raam-ne jakkham khurkeyaa*
ram-erg wound.m.sg scratch.perf.m.sg
'Ram scratched the wound.'

On the other hand, this semantic factor is unavailable for optional ergative subject taking unergatives 'to laugh', 'to dance', where the object cannot be demarcated as an entity separate from the event referred to by the unergative predicate. Evidence for this proposal is provided by (18) where it is possible only to individuate discrete or independent episodes of laughter, and not the entity 'laugh' itself. Thus, (19) without the use of the lexical item corresponding to 'instance' is infelicitous. Contrast this with (20) for the type I unergative 'spit', where use of the numeral refers to the plural number of spits and not to multiple episodes of spitting.

18. *jɔn /jɔn-ne caar vaarii hassii hasseyaa/hassii*
John.nom/John-erg four instance laugh.f.sg laugh.perf.m.sg/f.sg
'John laughed four times.'
19. *#jɔn /jɔn-ne caar hassii hasseyaa/hassii*
John.nom/John-erg four laugh.f.sg laugh.perf.m.sg/f.sg
'John laughed four times.'
20. *raam-ne caar thukkaã thukkiyaã*
ram-erg four spit.f.pl spit.perf.f.pl
'Ram spat four spits.'

To sum up, we have shown that while unergatives in Punjabi seem to fit into a structural class with respect to certain diagnostics, they show semantic variation. We proposed that it is this semantic variation that translates into a different structure for type I unergatives, resulting in an obligatory ergative subject. This class of unergatives therefore seems closer to transitive verbs which must occur with an ergative subject. This is in contrast with the structure for type II unergatives, which take an ergative subject optionally.

References

Bhatt, R. 2003 Topics in the Syntax of the Modern Indo-Aryan Languages. Causativization. Handout

Bobaljik, J. D. 1993. Eragtivity and Ergative Unergatives. In MIT Working Papers in Linguistics: Papers on Case and Agreement Phillips, Bobaljik, J., Phillips, C. (Eds.), 19: 45-88

Laka, I. 1993.Unergatives that assign ergative, unaccusative that assigns accusative. In MIT Working Papers in Linguistics: Papers on Case and Agreement Bobaljik, J., Phillips, C. (Eds.), 149–172

Mahajan, A. K. 2012. Ergatives, Antipassives and the Overt Light v in Hindi. Lingua 122. 204–214

Mohanan, T. 1994. *Argument Structure in Hindi*. CSLI Dissertations in Linguistics, Stanford

Preminger, O. 2012. The absence of an implicit object in unergatives: New and Old Evidence from Basque. Lingua, 122(3), 278-288

Richa 2008.Unaccusativity, Unergativity and the Causative Alternation in Hindi: A Minimalist Analysis. Doctoral Dissertation, Jawaharlal Nehru University, India

Linguistic structures in ESL learners' diaries

Varalakshmi Chaudhry
Guest Faculty, LEC
JNU, New Delhi
varalakshmichaudhry@yahoo.co.in

1 Introduction

The basic argument in the paper is that Diary Writing is the ideal task to improve the English language proficiency of learners. A study of the learner diaries gives us an insight into the mind of English as Second Language (ESL) learner and an idea of the cognitive processes involved while writing the diary. Diary writing task is a useful one to understand learner cognition (Ellis, 2005). The role of input-processing and output-processing in SLA (Van Patten, 2004) are undeniable. There are very few studies (Chaudhry, 2014) in the area of the cognitive processes triggered during task performance and the input/output processing that it brings in its wake to facilitate SLA.

2 Teaching-Learning Situation

The teaching-learning situation for the study is the Linguistic Empowerment Cell (LEC) of Jawaharlal Nehru University (JNU), New Delhi. LEC offers Basic Communication Skills course for students of different courses in the various Schools of JNU. The purpose of the course is to develop the communication skills of the students in English; so that they can comprehend the lectures in their regular classes and communicate the same in writing when required.

3 The Study

For the present study, ten students of such a course are given the task of diary writing. The students were asked to maintain a diary for 4 weeks. The diary would have the following details: Activities that they are involved in on a particular day; Thoughts that came into their mind during the day; and the diet details of that particular day. Although all the ten students couldn't maintain the diary for all the 4 weeks, the data collected was sufficient to get an idea about their linguistic repertoire.

4 Data Analysis

Subject	WK	Days	No. of pages	No. of sentences
S1	1	7	Typed: D1 – 1, D2 - .5, D3 - .5, D4 - .5,D5 – 1, D6 - .5, D7 – 1	D1 – 57, D2 – 36, D3 – 28, D4 – 31, D5 – 62, D6 – 37, D7 – 40
	2	7	Typed: D1 – 1, D2 – 1, D3 – 1, D4 – 1, D5 – 1, D6 – 1, D7 - .5	D1 – 71, D2 – 57, D3 – 85, D4 – 76, D5 – 90, D6 – 60, D7 – 41
	3	7	Typed: D1 – 1, D2 – 1, D3 – 1, D4 – 1, D5 – 1, D6 – 1, D7 – 1.5	D1 – 65, D2 – 62, D3 – 83, D4 – 75, D5 – 78, D6 – 69, D7 – 102
	4	5	Typed: D1 – 1, D2 - .5, D3 – 1, D4 – 1, D5 - 1	D1 – 78, D2 – 30, D3 – 68, D4 – 53, D5 – 55
S2	1	4	Handwritten: D1 – 2, D2 – 1.5, D3 – 1.5, D4 – 1.5	D1 – 33, D2 – 22, D3 – 22, D4 – 17
S3	1	7	Handwritten: D1 – 2, D2 – 2, D3 – 2, D4 – 1.5, D5 – 2, D6 – 2, D7 - 2	D1 – 29, D2 – 31, D3 – 31, D4 – 22, D5 – 33, D6 – 19, D7 – 27
S4	1	5	Handwritten: D1 – 2, D2 – 2, D3 – 1, D4 – 1.5, D5 – 2	D1 – 50, D2 – 56, D3 – 30, D4 – 51, D5 – 58
	2	6	Handwritten: D1 & D2 – 2, D3 - .5, D4 – 1.5, D5 - .5, D6 - .5	D1 – 55, D2 – 4, D3 – 38, D4 – 10, D5 – 14, D6 – 14
	3	7	Handwritten: D1 - .5, D2 - .5, D3 - .5 D4 – 2, D5 – 2, D6 & D7 - 2	D1 – 14, D2 – 5, D3 – 20, D4 – 46, D5 – 50, D6 & D7 – 41
	4	1	D1 – 2	D1 – 51
S5	1	6	Handwritten: D1 – 2, D2 – 2, D3 – 2, D4 – 2, D5 – 2, D6 - 2	D1 – 111, D2 – 98, D3 – 106, D4 – 102, D5 –96, D6 – 104
S6	1	5	Handwritten: D1 – 2, D2 – 2, D3 – 2, D4 – 1.5, D5 - 2	D1 - 90, D2 – 87, D3 – 69, D4 – 45, D5 - 63
	2	5	Handwritten: D1 – 2, D2 – 2, D3 – 1, D4 – 2, D5- 2	D1 – 57, D2 – 65, D3 – 48, D4 – 56, D5 – 56
	3	3	Handwritten: D1 – 54, D2 – 58, D3 – 2	D1 – 54, D2 – 58, D3 – 59
S7	1	7	Handwritten: D1 – 1, D2 – 1, D3 – 1, D4 – 1, D5 – 1, D6 – 1.5, D7 - 2	D1 – 28, D2 – 32, D3 – 46, D4 – 39, D5 – 36, D6 – 51, D7 - 60
S8	1	6	Handwritten: D1 – 2, D2 – 2, D3 – 65 D4 – 2, D5 – 2, D6 - 2	D1 – 60, D2 – 73, D3 – 65, D4 – 61, D5 – 45, D6 – 44
S9	1	2	Handwritten: D1 – 3/4 page, D2 – 1	D1 – 25, D2 – 34
S10	1	2	Handwritten: D1 – 1, D2 - 1	D1 – 27, D2 – 24

Table 1: Quantitative Analysis

Sub	Connectors	Time indicators
S1	But, And, Because, So, Usually, Additionally, Fortunately, Instead of	Every Monday & Thursday, After that, Before, In between, Last Monday, Every year, When,
S2	And, Because, But , As usual	Today, After that, After sometime, After they went back, After prayer
S3	Then, Suddenly, And, But, Because, Finally, Now	At 7 am, For 5 hours, After that, Good morning, Good night, Today, For 7 hours, Evening tea
S4	So, And, Well, And how, But why, Hopefully, But,	After that, And then, And now, At 12.30 pm, Yesterday, Every Saturday, Half-an-hour, Today, Tomorrow,
S5	Because, So, And, First time,	Today, At 6.45am, 9 o' clock, But today, On 15 Sept 2014, Tuesday & Friday, 8 to 10am, Next day
S6	Though, On the way, But, And oh! So, Overall, Another matter, Well	By 13.45, After taking lunch, Meanwhile, Half an hour, Lunch time, Two hours, Good night, Morning,
S7	Moreover, But, Sometimes, But again, But for me, As, Again	Today, In the afternoon, At 4 pm/ 6 pm, then, After taking dinner, Now, At 10o' clock, In the afternoon
S8	From there, But, And, So, As, Because	Today, At 7:00 clock/ 6:50 am /7:45/ 3.25pm/ 4:20 pm, 5-10 minutes later, After that, Till 10o'clock, Then,
S9	Because, But, As	At 6.30, At 7 am, For ten minutes, After that, At the same time, Then,
S10	And, So	At Friday/Saturday, 19th September 2014, At 5 am/ 7:30 am/ 9.30 am

Table 2: Types of Linguistic Structures

The quantitative analysis clearly reveals the limited linguistic repertoire of the students. The simple sentence is resorted to for the majority of the diary-writing exercise. The simple sentences are supported by connectors and time indicators (as it is a diary). There is not much use of compound or complex sentence structures. With minimum linguistic competence, the students managed written performance with a minimum of errors.

Examples of linguistic structures used:

1. I woke up **and** directly did Shubuh prayer.
2. **Usually before go to bed** we played around with my cat Dana.
3. **Today** I made prawn butter sauce **and** vegetables stir.
4. **At first time in Delhi** my husband was nervous **but after a while** he is used to facing this habit.
5. **In between** we had lunch and did prayer.
6. **After that** I started the computer **and** checked my e-mail.
7. **After they went back till 2:00 o'clock** I was working with computer, **and** after 2:00 o' clock I did my launch **and till night** I was reading a book.
8. **When** I walking suddenly one girl stop her bycycle **and** said to me you looking very pretty.
9. Well, **after that** I went over to Bhabhi's **and** got my RCT completed.
10. **Today** I woke up at 6.45am it is too early specially for me **because** from past one month normally I woke up at 8 o'clock or 9 o' clock.

11. I did the morning tasks in a hurry **and** left the room for JNU.

12. **In the afternoon** I took my lunch.

13. **Today** I got up at 6.30 am **but** due to laziness I set up an alarm at 7:00 o' clock **as** I had to reach the Linguistic class.

14. **After that for ten minutes**, I talked to my friend about our plan for today.

15. I'm from Indonesia **and** in India I am stay in Munirka Budh Vihar Summan Vila.

The examples clearly indicate the use of time indicators and connectors in addition to the simple sentence structure.

5 Conclusion

Diary writing task is the best task to trigger the cognitive processes of a student in the process of collecting information for the diary. The more the learner thinks of creative content for the diary, the more there would be struggle to search for linguistic structures and hence there would be linguistic empowerment as well. Small-scale classroom-based action research studies by teachers would definitely give more insights into the kind of tasks that promote input-processing and subsequent SLA.

References

Chaudhry, Varalakshmi (2014) `Chunking in ESL Learner Data' in Binod Mishra & Prashant Mishra (eds) ``*Post colonial Pedagogical Issues: Strategies, Theories and Practices in English Language Teaching'* (pgs 169 – 188) New Delhi: Adhyayan ISBN: 978-81-8435-398-3

Ellis, R. (2005) *Analyzing Learner Language* London: *OUP*

Krashen, S.D. (1985) *The Input Hypothesis: Issues and Applications*. London: Longman

Prabhu, N.S (1987) *Second Language Pedagogy*, Oxford:

Skehan, P. (1998) *A Cognitive Approach to Language Learning*, Oxford: OUP.

Swain, M. (1985) "Communicative Competence: Some *Roles* of Comprehensible Input and Comprehensible *Output* in its Development" In S. Gass and C. Madden (eds) *Input in SLA*. Rowley, M. A: Newbury House.

Van Patten, B. (2004) Input Processing in SLA. In B. Van Patten (ed), *Processing Instruction: Theory, Research and Commentary*. New Jersey: Lawrence Erlbaum.

Generation of a Dialect Corpus in Khortha used in Jharkhand India: Some Empirical Observations and Theoretical Postulations

Niladri Sekahr Dash
Indian Statistical Institute, Kolkata
ns_dash@yahoo.com

Atul Aman
Indian Statistical Institute, Kolkata
atul.aman1@gmail.com

1 Introduction

This paper describes a process of collecting varied and moderately adequate amount of dialect data and linguistic information from an endangered dialect Khortha spoken in some districts of Jharkhand, India (Dangi 2012, Ohdar 2007). It also tries to address some empirical and theoretical issues involved in the method used to collect spontaneous speech data – in the form of the Khortha Dialect Corpus (KDC) – from the native Khortha speakers.

The KDC is developed not only with sets of basic lexical list and primary sentence types, but also with Free Discourse Text (FDT) that includes samples of speech occurring in various natural settings and contexts. Speeches are digitally recorded at different geoclimatic environments and sociocultural settings through face-to-face interviews with the Khortha speakers of different regions, age, sex, ethnicity, and occupation. Speech samples are collected to develop a multidimensional-cum-multidirectional corpus for the purpose of documentation and digitization of the dialect with reference to its culture, history, and heritage (Austin 2010). The methodology proposed here have strong functional relevance in preservation and promotion of the endangered indigenous Indian dialects for betterment of the dialect communities of the nation.

2 The Khortha Dialect Corpus (KDC)

Dialect corpus, based on amount and variety of data, is of two types: complete and incomplete. A dialect corpus is complete if data and information extracted from it is fit to portray a complete picture of the dialect used by the target community. On the other hand, a dialect corpus is incomplete if data and information obtained from it fail to project on the total panorama of the dialect but reflects maximally on the linguistic fabric of the dialect community (Austin 2006). Since the first type of dialect corpus is impossible to generate due to its constructional limitations, the second type is the best option and it can be generated by systematic use of the methods of corpus linguistics (Dash 2008).

Based on the nature of interference by the field workers, on the other hand, a dialect corpus may be of three types: monitored, guided, and free. While the first two types are mostly human-controlled, the last one is adequately free from it. Due to this feature, it includes all kinds of spoken interaction that are produced by native informants in their own ways without being controlled by field workers. The domains of such spoken discourses may be related to all possible spheres of life of informants and natural to all geoclimatic and sociocultural environments in which the community lives (Lehmann 2001). Since the goal of the KDC is to be maximally free from all kinds of interference provoked by field workers, it is imperative to say that the KDC has invariably tried to contain speech data generated in the third type.

The basic characteristic features of the KDC are many. It is free from all kinds external interference; large in amount of data; varied in text types and text genres; natural in setting, situation and background; spontaneous in actual text generation; uncontrolled in surreptitious expression; non-monitored by data collectors; non-interfered by language investigators or neighbouring stand-by observers; elaborate in narratives and expressions; synchronic as well as diachronic in spatio-temporal dimension; multi-directional in texture and orientation; largely representative in text coverage; and non-bias in text representation. That means, it covers almost all the aspects of life, living, and society of the Khortha dialect community with equal un-bias reflection on each and every aspect based on which the Khortha community will be able to establish and confirm its unique individual linguistic identity with regard to other sister dialects of the same geographical area as well as with other distant neighbouring dialects (Dash and Aman 2013).

3 Relevance of the KDC

The structure of the KDC – after it adds up some new issues and aspects in dialect data collection – adds a new dimension to the traditional dialectology. New techniques and strategies of corpus linguistics and language documentation are used to generate the KDC which is intend to be utilized in the following works besides being used as a corpus in regular tasks of traditional dialectology (Crowley 2007):

(a) Developing a representative KDC with speech data and dialect related information from the community.
(b) Storing corpus in digital form in an archive with extra- and intra-textual annotations.
(c) Processing KDC to make data and information ready for man and machine.
(d) Analyzing corpus to draw inferences to present the state and status of the dialect.
(e) Utilizing KDC for preserving and restoring the dialect, its culture, and its heritage.
(f) Developing text materials and resources for the promotion of the dialect and the community.
(g) Extracting data and information from the KDC for language planning and language revival.
(h) Making KDC and related information available for global use for promotion and preservation of the history and heritage of the dialect.

Since adequate amount of data and dialect-related information are successfully collected from native Khortha speakers from their indigenous geo-climatic settings and socio-cultural background through face-to-face interviews and recordings, it is believed that the goals stated above may be achieved if retrieved data and information are systematically analysed and interpreted (Bowern 2008).

4 4. Domains of the KDC

KDC contains speech samples elicited from different settings of spoken interaction. It, therefore, contains two vital components of text-type: imaginative text in the form of narrative text, and informative text in the form of dialogue. The form and composition of the KDC is realised from Table 1 which provides some ideas about the domains from where texts are elicited from the informants to develop the KDC.

Khortha Dialect Corpus (KDC)		
Imaginative Text	ballads, elegies, events, fables, fairy tales, folklore, folksongs, folktales, general stories, geography, ghost stories, history, idioms, legends, love stories, lullabies, mythology, oral stories, plays, poems, proverbs, puzzles, rhymes, riddles, songs, tales, etc.	Narrative Text
Informative Text	agriculture, ailments, business, commerce, cultivation, cults, culture, customs, environment, faiths, feasts, festivals, folk science, games, geography, health, history, hygiene, literature, nature, norms, norms, politics, practice, professions, relations, religion, rituals, social life, social rules, socialisation, sports, traditions, etc.	Dialogic Text
Informative Text	personal narrations, eye witness accounts, quarrels, instructions, reminiscences, conversations, arguments, mimicries, mediations, negotiations, onomatopoeias, etc.	Narrative Text

Table 1: Areas covered for data for DDC

The Table 1 shows that the KDC is made up with samples of narrative text and dialogic interaction that are obtained in a systematic manner following the rules and methods of corpus generation to make KDC maximally varied and widely representative of the dialect from which it is developed. Thus KDC becomes functionally suitable to address and satisfy

the requirements of all concerned who line up for data and information about the dialect and its users.

5 Questionnaire Used for the KDC

Whether language data is collected in written or in audio-digital format, a questionnaire is mandatory to guide fieldworkers in their tasks by keeping all their queries in front of the informants and retrieving the required information properly and efficiently (Abbi 2001). To develop KDC an elaborate questionnaire is prepared after discussions with some experts of the field. The questionnaire that has been used for this purpose consists of four major parts (Table 2).

Khortha Dialect Corpus (KDC)	
Part-I	Demographic information of informants
Part-II	1000 objective type questions
Part-III	25 subjective type questions
Part-IV	Additional data and information

Table 2: Questionnaire used for collecting KDC

Part-I contains questions relating to demographic information of Khortha informants like their name, age, gender, ethnicity, education, languages known, occupation, interests, family background, like or do not like his/her language, etc. Part-II contains one thousand 'one-word-answer' type questions relating to various words they use in their regular life (basic vocabulary). Part-III contains twenty-five subjective questions where the informants are triggered to reply questions in an elaborate fashion expressing their views and opinions about their life and society in their mother tongue; Part-IV contains texts relating to additional data and information the informants want to share with the team members with regard to their replies elicited against the questions. Questions of Part-III and Part-IV relate to Free Discourse Text (FDT) for better representation of attitude of the speakers with little or no interference from the data collector(s).

6 Selection of Informants

All informants selected for interview knew at least two languages: Khortha as their mother tongue, and Hindi as a second language. Initially it was decided to select 50 informants from different locations of the state. Later 100 native Khortha speakers within the age group of 50+ (both male and female) living in different districts of the state are selected. All the interviews are conducted at their place of residence, cultivating land, work place, local market, on the road, or at the community centre (near the temple) in the village.

7 Elicitation Techniques Used

Interviews are conducted on the previously prepared questionnaires thoughtfully designed to capture all possible shades and shadows of life and living of the Khortha speech community. Scheduled as well as analytical elicitation techniques (Samarin 1967: 108-120) are adopted as both the techniques have strong functional relevance in generation of the KDC. The use of the scheme (Fig. 1) helps to develop the KDC which is non-skewed, linguistically balanced, and culturally representative of the Khortha community.

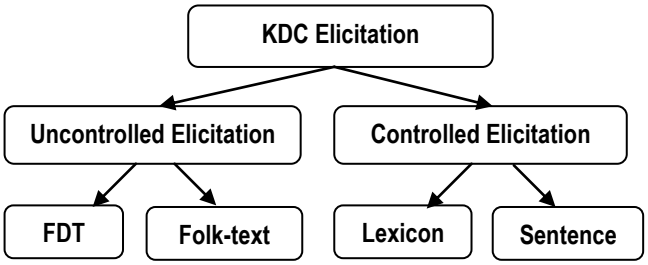


Fig. 1: Components of the digital dialect corpus

The diagram (Fig.1) shows that the KDC is made through two different methods: uncontrolled and controlled. The first method is invoked to elicit speech data from informant where there is no control in the inflow of the text generated by informants. In this case, the main task is to trigger informants to produce appropriate descriptive narratives relating to various aspects of their life, living, society, history, heritage, and culture, etc. The other part of the elicitation is to record the folk-text that constitutes songs, riddles, ballads, elegies, folktales, fables, old stories, anecdotes, lullabies, and other time-tested verbal texts transmitted through generations via oral transmission. This has been one of the most valuable components of the KDC, as it provides good scope to collect many antique words and old terms that are almost obsolete in the present form of the dialect (Sahu 2012).

The second strategy is a controlled one where a well-formed questionnaire with a list of thousand words and 500 sentences is used. This database is believed to be available in all natural languages (Swadesh 1955, Abbi 2001). Through continuous questioning, elucidation and interference conceptual equivalents of words and sentences are collected from the informants.

8 Composition of the KDC

Following the method stated above (Section 7) data of the following four types are collected from the native Khortha speakers:

(a) Basic vocabulary,
(b) Multiword expressions,
(c) Basic sentence types, and

(d) Free Discourse Speech

The role of data collector is minimised – limited to the act of a passive listener with conscious and careful alienation from on-going verbal interactions among the informants. The following list (Table 3) shows the number of hours of audio data collected from Khortha speakers in five linguistic surveys.

Survey	Data in MB	Data in Hours	Data in Words
S-I	185.41	3.40	28,621
S-II	349.58	6.45	50,229
S-III	177.43	2.25	21,924
S-IV	526.79	6.35	51,372
S-V	530.00	6.40	55,000
Total	1,755.21	25.25	2,071,46

Table 3: Total words and number of hours of data

It is hoped that this statistics will be meaningful in further research and in representing the dialect in a more faithful manner. The only limitation of this statistics is that it fails to show the distribution of the amount of data on different types of text. Effort is made to generate this information to be presented in the full report of the survey.

9 Utilization of the KDC

There are several reasons for developing the KDC as it has functional utility in various domains relating to life, language and culture of the Khortha dialect community. Because of its varied composition it is a good repository of rare and common, authentic and obsolete, general and specific use of linguistic items and elements. It is now available for retrieving most authentic speech data for understanding the general as well as special linguistic features and phenomena of the dialect. It can also be used for collecting data and information of the Khortha dialect to register its separate linguistic identity with regard to standard variety and other sister varieties.

Decision is taken to use KDC to develop general and special lexicon of Khortha; write descriptive grammar; produce text books and study materials; produce data for phonological, morphological, syntactic, and semantic analysis; compile reference dictionaries; and extract data and information for standardization of Khortha.

It is also planned to make the KDC available to others willing to use it to extract necessary data and information for other branches of social science like anthropology, ethnology, sociology, demography, sociolinguistics, psycholinguistics, ecolinguistics, history, culture studies, etc.

The KDC, after transcripted, may be annotated at phonological, grammatical, syntactic, semantic, and discourse level. Using customised annotation tools one can search into KDC to select and extract data considered relevant for particular investigation.

Finally, it is decided to use the KDC to develop a digital dialect archive for Khortha in which dialect data and linguistic resources will be preserved for future access.

10 Conclusion

The methods and strategies applied for developing the KDC has been useful in collecting required amount of speech data in the form of open text in a non-interference mode through direct involvement of informants. The success of this approach shapes up some theoretical postulations that challenge the traditional methods and processes used for dialect data collection in India. The value of this method may be further attested in schemes of large-scale dialect corpora generation works in India and abroad where people use digital tools and techniques to collect and store natural speech data from minority languages for future works relating to spoken language documentation, spoken text processing, speech technology, and dialect data digitization, etc.

In a country like India, the use of this method has high functional relevance for various sociolinguistic queries which may include a wide range of linguistic concepts, topics and issues that are considered relevant in study of dialects, dialect communities, culture, social psychology, language variation, ethnic diversities, ecology, diaspora, and other areas. Such queries are absolutely relevant for developing community profiles as well as postulating theories about dialects and their speakers.

References

Abbi, A. 2001. *A Manual of Linguistic Fieldwork and Structures of Indian Languages.* Muenchen: Lincom Europa.

Austin, P.K. 2006. Data and language documentation. In: J. Gippert, N. Himmelmann and U. Mosel (eds.) *Essentials of Language Documentation* (Trends in Linguistics. Studies and Monographs, 178), pp. 87-112. Berlin: Mouton de Gruyter.

Austin, P.K. 2010. Current issues in language documentation. *Language Documentation and Description*. 7(1): 12-33.

Bowern, C. 2008. *Linguistic Fieldwork: A Practical Guide.* New York: Palgrave Macmillan.

Crowley, T. 2007. *Field Linguistics: a Beginner's Guide.* Oxford: Oxford University Press.

Dangi, A.K. 2012. *Khortha Bhasha: Ek Parichay (Khortha Language: An Identity).* Jamshedpur, Jharkhand: Spardha Publications.

Dash, N.S. 2008. *Corpus Linguistics: An Introduction*. New Delhi: Pearson Education-Longman.

Dash, N.S. and A. Aman. 2013. An attempt towards documentation and preservation of the Khortha language of Jharkhand. Presented in *2nd National Seminar on Endangered and Lesser Known Languages* (ELKL-II), 23-24 October 2013, Dept. of Linguistics, Lucknow, India.

Lehmann, C. 2001. Language documentation: A program. In: Walter Bisang (ed.), *Aspects of typology and universals*, pp. 83-97. Berlin: Akademie Verlag.

Ohdar, B. 2007. *Khortha Bhasha Evam Sahitya: Udbhav Evam Vikash (Khortha Language and Literature: Origin and Growth).* Ramgarh, Jharkhand: Khortha Bhasha Sahitya Academy.

Sahu, C.R. 2012. *Encyclopedia of Indian Tribes.* New Delhi: Sarup Book Publishers.

Samarin, W.J. 1967. *Field Linguistics: A guide to Linguistic Field Work.* New York: Holt, Reinhart and Winston.

Swadesh, M. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics.* 21: 121-137.

Extending the core functionalities of Aṣṭādhyāyī 2.0

Oliver Hellwig
University of Düsseldorf
hellwig7@gmx.de

Wiebke Petersen
University of Düsseldorf
petersen@phil.uni-duesseldorf.de

1 Introduction

The paper describes new layers of linguistic annotation and explorative tools that were added to the project 'Aṣṭādhyāyī 2.0'. These additions make it possible to execute complex research queries in the digital version of Pāṇini's grammar with minimal knowledge both of Sanskrit and database query languages. In the project 'Aṣṭādhyāyī 2.0', we have developed a digital edition of Pāṇini's grammar of Sanskrit. Pāṇini introduced linguistic concepts, such as thematic roles, abstract derivation levels, rewrite rules, and pre-concepts of phonemes and morphemes, all of which are used intensively in contemporary Linguistics. In addition, the Aṣṭādhyāyī compresses the grammar rule system into text form by making use of inheritance structures, a sophisticated meta-language, and a marker system (Kiparsky 2009). Although these concepts and methods continue to be of highest scientific interest, the Aṣṭādhyāyī is rarely studied in modern formal Linguistics because readers who are not thoroughly acquainted with the Sanskrit grammatical tradition don't understand the terse Sanskrit text.

The research environment of 'Aṣṭādhyāyī 2.0' now opens up the content, the formal structure, and the encoding mechanisms of the Aṣṭādhyāyī to a wider scientific audience. We have built a web-based database edition with annotations on several linguistic levels.[6] While Petersen & Soubusta (2013) and Petersen & Hellwig (forthcoming) have dealt with the core database structure and the linguistic analysis, this paper describes the upper layers of annotation and introduces a customizable query mechanism for our database.

In Section 2, we will describe the annotation of *anuvṛtti* inheritance and of word-semantic concepts. Section 3 deals with the implementation of the search engine. Section 4 illustrates what kind of research problems can be tackled with the research environment Aṣṭādhyāyī 2.0.

[6] The web interface is accessible at http://panini.phil-fak.uni-duesseldorf.de/panini/.

2 Annotating semantic concepts and anuvṛttis

As described in Petersen & Soubusta (2013) and Petersen & Hellwig (forthcoming), we created the morphological and lexical annotation of the Aṣṭādhyāyī by closely following the analysis presented in Katre (1987). The same guidelines were adopted for the double checked annotation of the *anuvṛttis*. *Anuvṛtti* information is stored in a separate database table that records the number of a *sūtra* and the unique identifier of the component that is inherited in this *sūtra* according to Katre (1987). In several cases, components are not inherited over a continuous range of subsequent *sūtras*, but are missing in some of them. These rule blockings are discussed intensively in the Pāṇinian tradition and are sometimes controversial issues. So, adhering strictly to the *anuvṛttis* given in Katre (1987) was the most appropriate way of obtaining a homogeneous primary annotation. The *anuvṛtti* table contains a Boolean flag with which these blocked rules are marked in individual *sūtras*.

A modified version of the OpenCyc ontology[7] was used as the sense inventory for word semantic annotation. After having completed the annotation, we reduced the full ontology to an upper ontology by removing all branches that do not contain semantic concepts found in the Aṣṭādhyāyī. The remaining upper ontology has been reordered using Protégé[8], thereby performing simultaneously simplification of the concept hierarchy and domain adaptation. As could be expected, the structure of a modern Western ontology does not fit well the conceptual space of Indian texts. Animals, for instance, are ordered according to a Western scientific taxonomy in OpenCyc, while Indian texts frequently employ a conceptual subspace that resembles the structures found in the Amarakośa (Nair & Kulkarni 2010).[9] Another example for domain adaptation is the entry "writing" that is defined as "reading matter; anything expressed in letters of the alphabet" in the original ontology, and whose subclass "sacred text" contains the sibling classes "Veda" and "Mahābhārata". Obviously, Indian tradition would not insert the Vedas and the Mahābhārata under the same parent class (refer to the ordering described in Scharf (forthcoming)), and the overall parent class would not be considered to be a written, but rather an orally transmitted text. As the internal structure of the ontology influences the search results, this kind of domain adaptation is of highest importance for building a well usable query mechanism.

[7] http://www.cyc.com

[8] http://protege.stanford.edu

[9] Also refer to Hellwig (2014) for a comparative discussion of Western and Indian scientific taxonomies.

3 The search engine

Our data model is stored in a complex relational database that is tailored exactly to the research questions we are interested in (Petersen & Soubusta 2013). However, this database may be difficult to query for researchers without deeper knowledge of SQL. As systems are more likely to be used if their functionality is easy to understand (Davis, 1989), we decided to construct a simple query language that triggers the corresponding SQL-statements. This query language abstracts from the underlying relational database and reduces user queries to structured concatenations of keyword-argument pairs in the form of KEY(ARG). Arguments are either from fixed ranges of allowable values for a given keyword (e.g., gen(itive) or dat(ive) for the keyword "case"), or plain text, in which case the wildcards '*' (zero or any number of symbols) and '?' (exactly one symbol) are allowed. The logical operators AND, OR, and NOT are used to connect keyword-argument pairs. Users may employ brackets to assign priority to sub-clauses.

In order to allow the formulation of queries that are not supposed to be resolved on the level of a full *sūtra*, but on the level of a single component (which is usually a word), a special bracket type ('<','>') is reserved to bundle key-argument terms referring to one *sūtra*. For example, while 'case:dat AND num:pl' searches for all *sūtras* in which a dative component and a plural component occur, '<case:dat AND num:pl >' searches for all *sūtras* in which a component occurs that is a dative plural form.

Additionally, the search can be restricted to ranges of the Aṣṭādhyāyī by using the key term "snr" (*sūtra* number). 'snr(1-2)', for example, restricts the search to the first two books. More fine-grained selections are possible as well: 'snr(1.3-1.4)' searches in the third and fourth chapter of the first book, 'snr(1.2.4- 1.2.21)' restricts the search to the fourth to 21st *sutra* of the second chapter of the first book.

4 Research applications

Aṣṭādhyāyī 2.0 provides a customizable php web interface for viewing the linguistic annotation of Pāṇini's grammar. However, the higher level annotations and the query engine make it possible to handle complex research driven queries that go beyond the capacities of databases usually created in the Humanities:

- Combining query criteria: Our query language makes it easy to combine criteria from different annotation levels into one query statement. One example is Pāṇini's use of noun cases as metalinguistic markers. Using our query engine, one can search for lexemes in certain nominal

cases ("all instances of the word '*cu*' ('palatal') in the genitive case"), which are frequently constituents of rewrite rules.

- Querying semantic concepts: Aṣṭādhyāyī 2.0 started from the idea of making Pāṇini's grammar accessible to researchers with a limited knowledge of Sanskrit. Searching for semantic concepts is, therefore, an important step towards achieving this aim. Researchers can search for passages that contain particular semantic concepts or combinations of them ("all *sūtras* that contain a word meaning 'horse' or 'cow'") or even classes of semantic concepts ("all *sūtras* that contain one of the *pratyāhāras*"). Hellwig & Petersen (forthcoming) present a first case study for such queries.
- Making *anuvṛttis* visible: If desired, queries include the complete linguistic annotation of inherited elements for each *sūtra*. When searching for the concept *guṇa*, for example, the engine will retrieve, among others, *sūtra* 1.1.4 (*na dhātulope ārdhadhātuke*) that inherits the term *guṇa* from 1.1.3.

References

Davis, F. (1989), 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology', MIS Quarterly 13(3), 319-340.

Hellwig, O. (2014), 'Materialgruppen in der indischen Alchemie', Zeitschrift der Deutschen Morgenländischen Gesellschaft 164(2), 451-468.

Hellwig, O. & Petersen, W. (forthcoming), What's Pāṇini got to do with it? The use of *gaṇa*-headers from the Aṣṭādhyāyī in Sanskrit literature from the perspective of Corpus Linguistics, in 'Proceedings of the WCS 2015'.

Katre, S. M. (1987), Aṣṭādhyāyī of Pāṇini, University of Texas Press, Austin TX.

Kiparsky, P. (2009), On the architecture of Pāṇini's grammar, in G. Huet, A. Kulkarni & P. Scharf, eds, 'Sanskrit Computational Linguistics', Vol. 5402 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 33-94.

Nair, S. & Kulkarni, A. (2010), The Knowledge Structure in Amarakośa, in G. Jha, ed., 'Sanskrit Computational Linguistics', Springer Berlin / Heidelberg, pp. 173-189.

Petersen, W. & Hellwig, O. (forthcoming), Annotating and analyzing the Aṣṭādhyāyī, in 'Proceedings of CILC 2014'.

Petersen, W. & Soubusta, S. (2013), Structure and implementation of a digital edition of the Aṣṭādhyāyī, in M. Kulkarni, ed., 'Recent Researches in Sanskrit Computational Linguistics', D.K. Printworld, pp. 84-103.

Scharf, P. (forthcoming), Providing access to manuscripts in the digital age, in 'Writing the East: History and New Technologies in the Study of Asian Manuscript Traditions', Schoenberg Center for Electronic Text and Imaging, Philadelphia.

Frame Semantic Annotation of Sanskrit Texts

Oliver Hellwig
SFB 991, University of Düsseldorf
hellwig7@gmx.de

Sven Sellmer
Adam Mickiewicz University, Poznań
sven.sellmer@amu.edu.pl

1 Motivation and previous research

The paper describes an interface for annotating Sanskrit texts with FrameNet (FN) frames, and reports first experimental results of building an automatic labeler for Sanskrit FN frames. FrameNet is an established resource for frame semantic annotation of English (Fillmore et al. 2003). Each frame represents a situation and its typical participants. The situation itself is evoked by a target word in a text, and the participants are realized as core ("necessary") or non-core roles. Due to its broad coverage of typical situations described in texts, it is an obvious approach to port the English FN to other languages. Sanskrit is an especially interesting test case for this kind of domain adaptation, because its morphological, syntactical, and conceptual structures differ strongly from those of English, and because it has a large, well understood literature.

An overview of recent research in automatic Frame semantic annotation (FSA) is given in Das et al. (2014). As frequently in NLP, in-depth studies, resources, and working annotation pipelines for FSA are almost exclusively available for English. Problems that arise when FN is applied to non-standard English texts with an ancient Indian background are described in Reiter et al. (2010, 2011). A hybrid learning algorithm that annotates Sanskrit texts with valence based structures from a specialized semantic inventory has been applied to medieval alchemical texts in Sanskrit (Hellwig 2009a).

This paper is intended as a first explorative study of how FrameNet can be adapted to the conceptual space of Sanskrit. It discusses how Sanskrit texts can be annotated with semantic frames (Section 2), and how a statistical labeler for FSA can be designed for Sanskrit (Section 3). The concluding section 4 sketches some applications of frame semantics for Sanskrit philology and the cultural and historical study of Sanskrit literature.

2 The annotation interface

Frame annotation needs clearly demarcated target words to which frames and their roles are attached. Most of the tools used for frame annotation of Western languages split sentences by using white spaces, and present each of the resulting substrings of a sentence as a possible target for annotation. Apart from the fact that the notion of a sentence is not well defined in Sanskrit, this approach will fail due to *saṃdhi* and the intensive compound formation. Therefore, we propose to add the frame semantic annotation directly to the layer of lexicographic and morphological analysis that is generated by the program SanskritTagger (Hellwig 2009b, 2010). Because the tagger splits every surface string into its lexical components, roles and targets of a frame can be connected with their exact correspondences. In this way, the morphological and lexicographic information are preserved for later application of machine learning methods.

We have designed a `php` based web annotation interface that meets the special requirements of Sanskrit, including the non-ASCII letter set and the coordination between surface strings, their lexico-morphological analyses, and the frame annotations attached to them (refer to Figure 1). Annotation in our interface is done by (randomly) selecting a sentence with the desired target word, selecting a frame, and dragging its roles in the correct slots of the lexicographic analysis. The resulting annotations and the name of the annotator are stored in a relational database. In this way, our web-based UI allows collaborative, multi-user annotation with little administrative overhead for synchronizing the annotations.

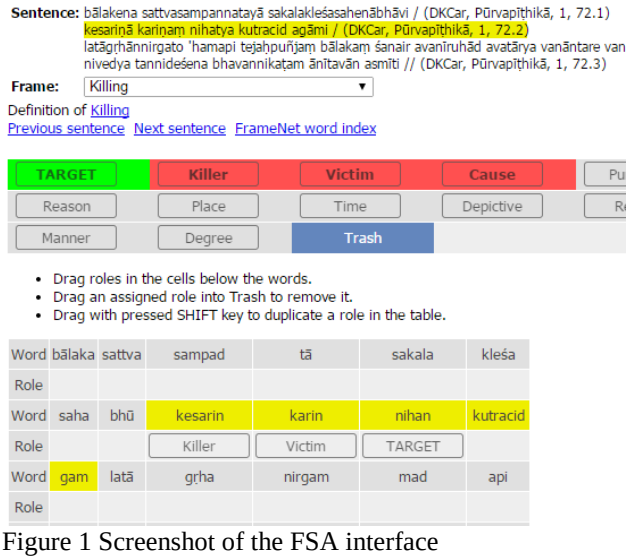


Figure 1 Screenshot of the FSA interface

3 Building an automatic labeler

Building a full system for automatic frame semantic labeling (FSL), as, for example, described in Das et al. (2014), is beyond the scope of this paper. Instead, we want to show that FSL is in principle possible for

a morphologically rich language such as Sanskrit, and that it can be implemented with comparatively few annotated data using "out of the box" algorithms.

To build up an initial database, a single annotator annotated 232 Sanskrit sentences with different frames. 101 of these sentences were annotated with the frame Education_teaching. This frame describes situations in which "a Student comes to learn either about a Subject; a Skill; a Precept; or a Fact as a result of instruction by a Teacher".[10] Typical target words for evoking this frame are *adhī* ("to learn"), *adhyayana* ("learning") or *anuśās* ("to instruct"). So, the Śloka Mahābhārata, 1, 1, 64.1 has been annotated as *idam*$_{\text{Subject}}$ *dvaipāyanaḥ*$_{\text{Teacher}}$ *pūrvam*$_{\text{Time}}$ *putram*$_{\text{Student}}$ *adhyāpayac*$_{\text{TARGET}}$*chukam*$_{\text{Student}}$.

The basic features used for labeling consist of (1) the lemma of each word, (2) its case and number, if it is declinable, and (3) a Boolean flag indicating if it is a verb. (1) - (3) are the seed features f_s that are created for each word in a Śloka. The final set of features F_i for a word at position i contains the seed features f_{si} of the word itself and of the words that are found in a context of size m around the word; so $F_i = \bigcup_{i+m}^{i-m} f_{sj}$. We used Conditional Random Fields (CRF, Lafferty et al. (2001)) as the learning algorithm, because they combine sequential decoding with the ability to handle high-dimensional feature spaces of nominal variables.

The evaluation was restricted to the annotated core roles of the frame Education_teaching, because there are not enough training data for its peripheral roles. It should, however, be emphasized that the labeler just rejected all non-core roles and produced no false positives in this way. The first section of Table 1 ("Without word semantics") shows the results of a leave-one-sentence-out cross-validation without word semantic information. For this sake, one of the 232 sentences was left out as a test case, while the remaining 231 sentences were used for training the labeler. This split was repeated for all 232 sentences, and the results of each test were concatenated to create the final evaluation set. While precision and recall for the target words and the precision for the role Subject (most frequently the Veda(s)) are acceptable, classification of the remaining core roles suffers from low recall.

The SanskritTagger database also contains word semantic annotation and information (Hellwig 2012). To generalize from the lexical level, we added word semantic information as the fourth feature to f_{si}. Results of this setting are printed in the middle section of Table 1 ("With word semantics"). Most remarkable are the low recall rates for the roles "Student" and "Teacher". Error analysis reveals that these roles are frequently occupied by infrequent personal names which the labeler has problems to learn. Therefore, we added referenced word semantic parent concepts up to a level of 2 to the seed features. In this way, the parent concept "man" can, for example, be added to the set of seed features of a personal name. Repeating the cross-validation with the parent semantic concepts produced the third section in Table 1 ("With word semantics and parent concepts"). While the F score for Student has slightly risen, the rates for Teacher remain unchanged. Sentences such as Bhāgavatapurāṇa, 3, 4, 20.1 (*ārādhitapādatīrthād*$_{\text{Teacher}}$ *adhītatattvātma-vibodhamārgaḥ*) are certainly responsible for the low recall rates. This finding emphasizes the need for large-scale annotation and for the abstraction from the lexicographic layer.

Role	Precision	Recall	F score	TP	FP	FN
	Without word semantics					
Student	53.85	6.93	12.28	7	6	94
Subject	71.74	34.74	46.81	33	13	62
TARGET	91.18	57.94	70.86	62	6	45
Teacher	80	10	17.78	4	1	36
	With word semantics					
Student	54.17	12.87	20.8	13	11	88
Subject	72.34	35.79	47.89	34	13	61
TARGET	91.67	61.68	73.74	66	6	41
Teacher	80	10	17.78	4	1	36
	With word semantics and parent concepts					
Student	63.64	13.86	22.76	14	8	87
Subject	77.27	35.79	48.92	34	10	61
TARGET	92.86	60.75	73.45	65	5	42
Teacher	80	10	17.78	4	1	36

Table 1: Results for labeling target words and core roles of the frame Education_teaching using CRF; leave one out cross-validation. TP: true positives, FP: false positives, FN: false negatives

4 Perspectives

Judging from the results reported in Section 3, designing a reliable FSL for Sanskrit texts is possible, but requires in-depth research into appropriate feature sets and classification methods. Most importantly, the comparatively free word order, the unclear notion of sentence boundaries, the rich morphology and the extremely diverse lexicography of Sanskrit need special attention. In addition, annotation guidelines have to be developed, and a multi-annotator approach must be employed for large-scale annotation.

Concerning the application of an FSL system, detecting and analyzing high frequency linguistic patterns in increasingly large digital corpora are

[10] 1Source of definition: https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Education_teaching

central challenges for current philology. FSL systems can be helpful in exploring word semantics when a sufficient number of annotations for a target word is available; consider, for instance, the use of the verb *han* and its prefixed forms over different domains such as descriptions of war ("to kill"), in medicine ("to cure a disease") and in alchemical texts ("to 'calcine' mercury"). However, their most important field of application is certainly the detection and the structured evaluation of frequent, stereotyped statements that can be used to reconstruct the mental history of ancient India.

References

Das, D., Chen, D., Martins, A. F. T., Schneider, N. & Smith, N. A. (2014), "Frame-semantic parsing". *Computational Linguistics* 40(1), 9-56.

Fillmore, C. J., Johnson, C. R. & Petruck, M. R. (2003), "Background to FrameNet". *International Journal of Lexicography* 16(3), 235-250.

Hellwig, O. (2009a), *Quecksilber und Chronometrie. Eine computerphilologische Analyse der mittelalterlichen indischen Alchemie*. Shaker Verlag, Aachen.

Hellwig, O. (2009b), "SanskritTagger, a stochastic lexical and POS tagger for Sanskrit". In G. Huet, A. Kulkarni & P. Scharf (eds) *Sanskrit Computational Linguistics. First and Second International Symposia*. Berlin: Springer Verlag.

Hellwig, O. (2010), "Performance of a lexical and POS tagger for Sanskrit". In G. Jha (ed.) *Proceedings of the Fourth International Sanskrit Computational Linguistics Symposium*. Berlin: Springer Verlag.

Hellwig, O. (2012), "ratha = 'warrior' or 'chariot'? Computational Approaches to Polysemy in Sanskrit". To appear in: *Proceedings of World Sanskrit Conference 2012*.

Lafferty, J. D., McCallum, A. & Pereira, F. C. N. (2001), "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". In *Proceedings of the Eighteenth International Conference on Machine Learning*.

Reiter, N., Hellwig, O. & Frank, A. (2011), "Semiautomatic semantic analysis of rituals" In Felder, E., Müller, M., Vogel, F. (eds.) *Korpuspragmatik: Thematische Korpora als Basis diskurslinguistischer Analysen*. Berlin: de Gruyter.

Reiter, N., Hellwig, O., Mishra, A., Frank, A. & Burkhardt, J. (2010), "Using NLP Methods for the Analysis of Rituals". In *Proceedings of the 7th International Conference on Language Resources and Evaluation*.

"Serial Verbs" in South Asian languages: Compound Verbs or Pseudo-Coordination?

Hans Henrich Hock

hhock@illinois.edu

Steever (1988) has introduced to South Asian linguistics the notion "Serial Verb" — a verb + verb construction in which the two verbs agree in person and number features, as in [1]. According to Steever, the first verb is syntactically subordinate, and agreement with the second verb is morphological; only the second verb is syntactically finite and hence licensed to be marked for person and number. Under this perspective, Serial Verbs can be considered special variants of the common South Asia Compound Verbs, except that the Pole is finite and agrees with the Vector.

[1]	celvēm	allēm
	go.NONPST.1PL	not.be.1PL
	'We will not go.'	

Steever further shows that while in much of Dravidian, Serial Verb structures are lost, Konda preserves a (possible) intermediate phase, with a truncated version of the agreement marker, as in [2].

[2]	1 SG.	*vā-t-a*	*suṛ-t-a*
		come.PST.TPE	see.PST.1SG
		'I came and saw'	
	PL	*vā-t-a*	*suṛ-t-a**p***
		come.PST.TPE	see.PST.1PL(EX)
		'We came and saw'	
	2 SG	*vā-t-i*	*suṛ-t-i*
		come.PST.TPE	see.PST.2SG
		'You came and saw'	
	PL	*vā-t-i*	*suṛ-t-i**der***
		come.PST.TPE	see.PST.2PL
		'You PL came and saw'	

I show that similar structures with Serial Verbs are found in other South Asian languages. In fact, all major language families have at least some members with Serial-Verb constructions. While in Munda languages, these might be attributed to Dravidian influence (Anderson 2003), in languages such as Gujarati ([3]) such an account is unlikely, and similar structures in Vedic Sanskrit [4] go back to Proto-Indo-European (Hock 2012).

[3]	*av-uṁ*	*ch-uṁ*
	come-1SG	be.PRS-1SG
	'I come'	

[4] a. *bhákṣa* + *éhi*
food.VOC.SG come.IMPV.2SG
mā + *á́ viśa* (TS 3.2.5.1)
I.ACC.CLIT enter.IMPV.2SG
'Come, food, enter me.'

b. *éta* + *u nv*
come.IMPV.2PL PCL
índraṁ *stávāma* (RV 8.24.19)
Indra.ACC.SG.M praise.SBJV.1PL
'Come now, let us praise Indra.'

Beyond demonstrating the existence of such structures outside Dravidian I focus on the variations in agreement that can be observed and propose that they can be accounted for in terms of a family of constraints. These range from full morphophonological agreement, as in [1]), to abstract agreement in terms of morphosyntactic features, but not necessarily phonological realization, as in [5], to attenuated agreement, as in [4], where the two verbs agree in modality but not necessarily in person-number features.

[5] *irakku* *vāreṉ*
beg.NPST.1SG come.NEG.1SG
'I do not come to beg.' (OTamil)

I further argue against the view that Serial Verbs are a subtype of Compound Verbs. While in the more common structures of the type [1], [3], and [5] the second verb can be considered the head of the construction, just as the Vector in Compound Verbs, and the first verb can be viewed as dependent on it, in Serial-Verb structures with "Balance Verbs" such as [6], there is no relation of subordination.

[6] *biba injo* *ōti*
marriage house.LOC take.CONV
samgiri *uṭar*
provisions drink.PST.3PL
ticar
eat.PST.3PL (Pengo)
'They ate-drank, i.e. consumed, the provisions brought to the marriage house.'

Moreover, accentuation in the early Vedic language shows that the relation between the two verbs is one of coordination, rather subordination; it is only in late Vedic that the second verb shows dependent-clause accentuation; see Hock 2012.

Under the circumstances, the Serial Verb construction is best accounted for as a special case of what is called Pseudo-Coordination in reference to English examples such as [7]; see Ross 2014. The major difference is that Pseudo-Coordination is defined as involving an original overt conjunction; Serial Verbs seem to have covert conjunction. Beyond that, however, both types of construction involve grammaticalization to begin with, and further grammaticalization can lead to restrictions on agreement — consider the examples in [2] for Serial Verbs and those in [8] for Pseudo-Coordination (Ross 2014). In most dialects examined by Ross, the latter require that the second verb be identical to the infinitive, but the construction as a whole requires identity in agreement features; not meeting both requirements makes [8c] ungrammatical, while [8d] is more acceptable, presumably because the "infinitive constraint" is stronger.

[7] a. What did he come and get?
b. Try and be good

[8] a. Try and be good
b. We try and be good
c. *He tries and is good
d. ??He tries and be good

References

Anderson, Gregory D. S. 2003. Dravidian influence on Munda. *International Journal of Dravidian Linguistics* 32(1): 27–48.

Hock, Hans Henrich. 2012. Come and Get It: The Indo-European background of the Vedic *éta ... stávāma* construction. In: *Proceedings of the 24th Annual UCLA Indo-European Conference*, ed. by Stephanie W. Jamison, H. Craig Melchert, and Brent Vine, 47-66. Bremen: Ute Hempen Verlag,

Ross, Daniel. 2014. Between coordination and subordination: Typological, structural and diachronic perspectives on Pseudocoordination. http://publish.illinois.edu/djross3/?attachment_id=162

Steever, Sanford B. 1988. *The serial verb formation in the Dravidian languages*. Delhi: Motilal Banarsidass.

Romanisation of Indian languages: a diachronic analysis of its failure

Arvind Iyengar
University of New England, Australia
arvind.iyengar@une.edu.au

1 Introduction

This paper will cover the rich but understudied history of the usage of the Roman script (ABCD…XYZ) for the languages of the Indian Subcontinent. The romanisation of languages in various non-European multilingual regions, such as East and South-East Asia, Africa and the Middle East, has been relatively well-documented and studied. However, any comprehensive diachronic analysis of the—ultimately unsuccessful—attempts at romanising languages in India has yet to appear. This paper is an initial effort in that direction.

This paper comprises two sections. The first section, spanning the years 1616 to 1947, will first deal with the initial efforts by Europeans in India to transcribe Indian languages in the Roman script. It will then describe the subsequent attempts by various European and Indian proponents to introduce a uniform romanisation system throughout India, and their eventual failure. The second section will seek to answer various questions that arise during the preceding historical analysis, based on the Framework for the Comparative Analysis of Romanisation (Aytürk 2010a, 2010b).

Since this paper largely deals with the period prior to the Partition of British India into modern-day India and Pakistan in 1947, the term 'India' in this paper shall mean the entire Indian Subcontinent.

2 From missionary grammars to mass romanisation (1616-1947)

The history of romanisation for Indian languages goes at least as far back as the seventeenth century. The initial proponents of the Roman script were European Christian missionaries in India, who made use of Roman for writing grammars of and translating Biblical scriptures into Indian languages. Probably the earliest such romanised work in an Indian language was that of the English Jesuit Thomas Stephens, who was stationed in the Portuguese colony of Goa in western India. Stephens' *Christa Purana* (1616), an epic poem on the life of Jesus Christ, was written in a mix of Marathi and Konkani, and printed entirely in the Roman script. The reasons for Stephens choosing the Roman script were the unavailability of Devanagari type and the Roman script ostensibly being known both to the Indians and the Portuguese (Pereira 1992; SarDessai 2000).

With the increase in trade and contact between Europe and India, the seventeenth century saw an increased interest in Indian languages among European Christian missionaries as well as merchants. The orthographic conventions used for transcribing Indian languages in Roman were often influenced by the writer's first language (e.g. German, Italian) or the language in which the work was written (e.g. Latin).

Following the establishment of the first British colonies in India in 1757, British authors started compiling works and textbooks on Indian languages not just out of missionary, scholarly or commercial interest, but primarily "to learn the language of the country they were colonising and . . . to figure out how to do this" (Friedlander 2006).

Possibly the first—or at the very least, the best known—attempts at proposing a systematic transliteration of Indic and Arabic characters were those of John Borthwick Gilchrist (1796) and William Jones (1799).

Gilchrist put forward a regular—albeit Anglicised—Roman transliteration for the letters of the Hindustani language. His aim was to enable British learners to attain a reasonably authentic pronunciation of Hindustani, without them having to learn the native scripts of the language.

Jones' target, on the other hand, was explicitly the orthography. His outlook was also broader, in that he proposed a uniform transliteration system for "Indian, Arabian and Persian" letters (1799, p. 175). Jones' proposal was thus focused on Sanskrit, Arabic and Persian, rather than Hindustani.

The 'Jonesian' and 'Gilchristian' systems became synonymous with *orthography* and *orthoepy*, namely with spelling and pronunciation respectively (Markham 1878). Both systems found their adherents, but neither system became commonly adopted by the general public, among whom erratic romanisation of Indian words continued to prevail.

The debate on whether and how to romanise Indian languages continued well into the nineteenth century. Notable figures in favour of blanket romanisation of major Indian languages included the Indologist Friedrich Max Müller, the Sanskritist Monier Monier-Williams, the bishop Robert Caldwell and the civil servant Sir Charles Edward Trevelyan. Those against such romanisation were the orientalists James Prinsep and John Tytler. Arguments for romanisation included the simplicity, linearity and consequently reduced cost of printing in Roman letters, and the convenience of using a uniform alphabet throughout British India. A covert aim of romanisation among some proponents was the eventual evangelisation of the masses (Yates,

Pearce, Trevelyan, & Thomas 1859 [1834]). Arguments against romanisation included the superior compactness of the Indic and Arabic scripts, and the inadequacy of the Roman script in representing phonemes peculiar to Indian languages.

By the 1850s, there were no less than twenty-two different romanisation systems for Indian languages (Max Müller 1854). In light of this, a trade-off romanisation scheme proposed by William Wilson Hunter, a Scottish historian and civil servant stationed in Bengal, was finally adopted by the British Indian government in 1870. Hunter reasoned that the Jonesian system contained too many diacritics for common usage. He therefore proposed a derivative system with a minimum number of diacritics as a compromise (Skrine 1901).

The official adoption of the Hunterian system, namely a trade-off with sub-optimal phonetic accuracy, likely led to the dwindling of calls for sweeping romanisation in India.

However, by the early twentieth century, with Indian self-rule looming large, the question of national language and with it, national script, were rekindled. Linguist Suniti Kumar Chatterji argued that on a pan-Indian level, "[t]he problem of the Babel of scripts in India presents itself . . . as being capable of a final solution only through an Indo-Roman script" (Chatterji 1935). However, with Roman being popularly perceived as the coloniser's script, Chatterji also foresaw widespread resistance to such romanisation.

The Second World War provided an unexpected arena for the romanisation of Indian languages, particularly Hindustani. War propaganda pamphlets distributed by the rebel Axis-aligned Indian National Army (INA) were commonly written in romanised Hindustani (Friedman n.d.). This was presumably done not only to reach out to the largest possible audience, but also to implicitly bridge the Hindu-Muslim divide that had grown increasingly bitter by the 1940s. The use of Devanagari by Hindus and Perso-Arabic by Muslims for writing Hindustani served as an explicit manifestation of this divide. The INA thus successfully harnessed the 'enemy' Roman script not just to increase readability, but also to symbolise communal harmony and present the picture of a united nation.

With the ultimate defeat of the INA by the Allied powers and the bloody Partition of British India along religious lines in 1947, the INA's attempt at mass romanisation in India, like most others before it, also faded into oblivion.

Today, the Hunterian system nominally remains the government-approved romanisation system in India, Pakistan and Bangladesh (UNGEGN 2006), but its strict usage is by and large confined to government cartographic publications.

3 Failure of romanisation: an analysis

On the basis of the successful romanisation of the Turkish language and the unsuccessful romanisation of Hebrew in the 1920s, Aytürk (2010a, 2010b) has put forward a Framework for the Comparative Analysis of Romanisation (FCAR), which seeks to explain—and possibly predict—the success or failure of a particular attempt at romanisation.

The FCAR includes a variety of technical-infrastructural as well as political-cultural factors that demonstrate the likelihood of romanisation being successful. The following technical-infrastructural factors work in favour of romanisation:

- complex existing script
- low mass literacy in the existing script
- existing script more expensive to print in than in Roman

Similarly, the following political-cultural factors all contribute to romanisation:

- authoritarian regime in favour of romanisation
- environment of political upheaval
- prior exposure to Christianity

However, in spite of the above conditions being fulfilled in the colonial Indian context, romanisation did not succeed.

I therefore argue that the historical failure of romanisation in India is not comprehensively explained by the FCAR, and that additional factors need to be taken into account in order to fully contextualise and comprehend this failure. I therefore propose that the following additional factors be included in the FCAR for purposes of this analysis:

- Number of stakeholders
- Use of language and script in administration
- Availability of typefaces in local script
- Number of languages in question
- Size of target population

In other words, the following factors also need to be considered in order to satisfactorily explain the failure of romanisation in India:

- The number of parties with mutually contradictory expectations from romanisation (simplicity versus accuracy)
- The British policy of using local languages (Hindustani, Bengali etc.) as official government languages, thereby inadvertently entrenching the association of a particular script with the language
- The availability of type in the various Indic scripts, with several having been initially

cast by the British themselves, thus inadvertently leading to their standardisation
- The task of romanising not one but several languages, with differing phonemic inventories
- A target population of roughly 400 million people – one-sixth of the world population at the time

References

Aytürk, İ. (2010a). "Romanisation in Comparative Perspective". *Journal of the Royal Asiatic Society, 20*(1), 1-9.

Aytürk, İ. (2010b). "Script Charisma in Hebrew and Turkish: A Comparative Framework". *Journal of World History, 21*(1), 97-130.

Bose, S. C. (2010). *Words of Freedom: Ideas of a Nation.* New Delhi: Penguin Books.

Caldwell, R. (1859). "On the Substitution of the Roman for the Indian Characters". *The Madras Journal of Literature and Science, IV*(7), 243-271.

Chatterji, S. K. (1935). "A Roman Alphabet for India". *University of Calcutta: Journal of the Department of Letters, XXVII*, 1-58.

Friedlander, P. G. (2006). "Hindustani Textbooks from the Raj". *Electronic Journal of Foreign Language Teaching, 3*(1), 39-56.

Friedman, H. A. (n.d.). *Axis and Allied Propaganda to Indian Troops.* Retrieved June 2, 2014, from Psywarrior: http://www.psywarrior.com/AxisPropIndia.html

Gilchrist, J. B. (1796). *A Grammar of the Hindoostanee Language.* Calcutta: Chronicle Press.

Markham, C. R. (1878). *A Memoir on the Indian Surveys.* London: W. H. Allen & Co.

Max Müller, F. (1854). *Proposals for a Missionary Alphabet.* London: A. & G. A. Spottiswoode.

Monier-Williams, M. (ed.). (1859). *Original Papers illustrating the History of the Application of the Roman Alphabet to the Languages of India.* London: Longman, Brown, Green, Longmans and Roberts.

Pereira, A. (1992). "Stephens, Fr. Thomas". In M. Lal (ed.), *Encyclopaedia of Indian Literature* (Vol. V, p. 4182). Delhi: Sahitya Akademi.

Royal Asiatic Society. (1896). *Transliteration Report.* London: Royal Asiatic Society.

SarDessai, M. (2000). *A History of Konkani Literature: From 1500 to 1992.* New Delhi: Sahitya Akademi.

Shipley Jones, A. M. (Ed.). (1799). *The works of Sir William Jones* (Vol. 1). London: G. G and J. Robinson, R. H. Evans.

Skrine, F. H. (1901). *Life of Sir William Wilson Hunter, K.C.S.I., M.A., LL.D., a Vice-President of the Royal Asiatic Society, etc.* London: Longmans, Green, and Co.

Trevelyan, C. E. (1859 [1834]). "Mr. C. E. Trevelyan's First Minute in support of Mr. Thompson's dictionary, and of the Application of the Roman Alphabet to Hindústání". In M. Monier-Williams (ed.), *Original Papers Illustrating the History of the Application of the Roman Alphabet to the Languages of India* (pp. 3-30). London: Longman, Brown, Green, Longmans and Roberts.

UNGEGN. (2006, October 9-11). *The Romanization of Toponyms in the Countries of South Asia.* Retrieved May 28, 2014, from Working Group on Romanization Systems: http://www.eki.ee/wgrs/wgr06_5.htm

Yates, W., Pearce, W. H., Trevelyan, C. E., and Thomas, J. (1859 [1834]). "Circular Letter addressed by the Originators of the General Application of the Roman Letters to the Languages of the East, to the Principal, Tutors, and Students, especially Students preparing for the Ministerial and Missionary Work, at the different colleges in the United States". In M. Monier-Williams (ed.), *Original Papers Illustrating the History of the Application of the Roman Alphabet to the Languages of India* (pp. 163-174). London: Longman, Brown, Green, Longmans and Roberts.

Conjunct Agreement: Patterns and role of the person feature

Gurmeet Kaur
Indian Institute of Technology, Delhi
gurmeetk88@gmail.com

The phenomenon of agreement between a predicate and conjoined phrase (ConjP) made up of two DPs manifests a huge amount of variation cross linguistically. Hindi-Urdu allows the predicate to agree either with the linearly closest conjunct DP (called closest conjunct agreement/CCA) or with the entire conjP (Bhatia 2011; Bhatia, Benmamoun and Polinsky 2009). Slovenian, on the other hand, allows three options- agreement with the closest conjunct, agreement with the highest conjunct or agreement with the full conjP (Marusic, Nevins and Badecker 2012; Marušic, Nevins and Saksida 2007). These works stress on the interaction between linear proximity and the computation of agreement. Employing novel data from Punjabi, an Indo Aryan language, I present a different pattern of conjunct agreement with predicative adjectives, stressing on the importance of person features of the conjunct DPs. Specifically, I demonstrate that while linear adjacency, resulting in closest conjunct agreement plays an important role in agreement relations when both the conjunct DPs in a conjoined phrase are 3rd person nouns, it is rendered irrelevant when at least one of the DPs is 1st/2nd person. CCA becomes unavailable in such instances.

In this paper, I focus on the nominative conjoined phrases that trigger agreement on the predicative adjective and present the emerging agreement patterns. We start with the cases where both the DPs are 3rd person. The presence/absence of CCA in these cases is determined by the number specification on the two 3rd person conjunct DPs. When both conjuncts have the same number value (singular or plural), default agreement can (mas.plural) obtain on the predicate, as shown in (1-2). When the two conjunct DPs differ in their number specifications, CCA obligatorily takes place (3-4).

1. mundaa te kuRi khote/jhalle ne/
   *khoti e
   boy.M.sg and girl.F.sg stupid.m.pl
   be.pres.3.pl/*stupid.f.sg be.pres.3.sg
   'The boy and girl are stupid.'
2. munde te kuRiyaan khote ne /
   ??khottiyaan ne
   boy.M.pl and girl.F.pl stupid.m.pl be.pres.3.pl
   /??stupid.f.pl be.pres.3.pl
3. munde te kuRi khotii e/
   ?khote ne
   boy.M.pl and girl.F.sg stupid.f.sg be.pres.3.sg/
   ?stupid.m.pl be.pres.3.pl
4. mundaa te kuRiyaan khottiyaan ne
   /?khote ne
   boy.M.sg and girl.F.pl stupid.f.pl be.pres.3.sg/
   ?stupid.m.pl be.pres.3.pl

Another layer of complexity is added to the pattern by employing conjoined phrases with at least one of the conjuncts being 1st/2nd person. In such cases, the predicate-auxiliary obligatorily agrees in person with the 1st/2nd person conjunct. Number and gender agreement takes place with the conjP, corresponding to mas.pl or fem.pl. Note that in (5)-(7) given below, the auxiliary agrees with the most marked value of person. Thus, in (5) and (6) agreement is triggered by the more marked 1st person feature. Between a 2nd and a 3rd person DP in (7), it is the former which is more marked and controls agreement. The agreement facts remain unchanged irrespective of the structural and linear position of the agreement controlling DP; see (8). Importantly, in all of the cases listed in (5-8), there is no CCA.

5. maiN te tuu khote aaN
   1.sg and 2.sg stupid.m.pl be.pres.1.pl
   'I and you are stupid.'
6. maiN te o khote aaN
   1.sg and 3.sg stupid.m.pl be.pres.1.pl
   'I and he are stupid.'
7. tuu te o khote o
   2.sg and 3.sg stupid.m.pl be.pres.2.pl
   'You and he are stupid.'
8. tuu te maiN khote aaN
   2.sg and 1.sg stupid.m.pl be.pres.1.pl
   'I and you are stupid.'

So far, we have seen that with 3rd person DPs in conjunction, the predicate agrees either with the ConjP or with the closest conjunct. In presence of atleast one 1st/2nd person DP in the conjoined phrase, the predicate agrees with the conjunct with the highest person feature (1>2>3; Silverstein 1976), followed by agreement with the ConjP. No CCA can obtain in these instances. To account for these agreement patterns presented in (1)-(8), I adopt the following structure in (9) for the ConP in Punjabi.

9.[$_{ConjP}$DP1[ConjDP2]]

Evidence for the higher position of DP1 is provided by tests like binding and extraposition (Bhatia 2011; Munn 1999 etc.). Consider (10), where the first conjunct can bind into the second one, but not vice-versa, as in (11), indicating that DP1 is structurally higher than DP2 in the conjoined phrase.

10. munde-ne har kuRii$_i$ te oddi$_i$ maa -
    nuũ vekhyaa
    boy-erg every girl and 3.gen.F.sg
    mother-acc see.perf
    'The boy saw every girl and her mother.'

11. *munde-ne oddi$_i$ maa te har kuRii-nuũ$_i$ vekhyaa
boy-erg 3.gen.F.sg mother and every girl-acc see.perf
'The boy saw her mother and every girl.'

Further evidence that the conjunction is asymmetrically structured with DP1 being higher comes from extraposition. The structure in (9) entails that the conjunction particle and DP2 form a constituent for movement, separate from DP1. Consider example (12), where extraposition of the second conjunct along with the conjunction particle is allowed. Movement of either the second conjunct without the conjunction particle, as in (13) or the first conjunct with the conjunction particle in (14) leads to ungrammaticality.

12. munde-ne ikk kapp t_i toRyaa [te ikk glaas]$_i$
boy-erg one cup break.perf.M.sg and one glass
'The boy broke a cup and a glass.'
13. *munde-ne ikk kapp te t_i toRyaa [ikk glaas]$_i$
boy-erg one cup and break.perf.M.sg one glass
14. *munde-ne t_i ikk glaas toRyaa [ikk kapp te]$_i$
boy-erg one glaas break.perf.M.sg one cup and

With the agreement facts and structure of ConjP been presented, I now move on to giving a possible analysis for conjunct agreement in Punjabi. I adopt a two-step approach to agreement with (Benmamoun, Bhatia, and Polinsky 2009 etc.) wherein agreement between a Probe and Goal for a particular phi feature and the subsequent copying of that feature value are two distinct steps. Further, I assume that the resolution of features at ConjP is obtained by ConjP inspecting the features of both its conjuncts and ending up with a certain value. I present the derivational steps for agreement in three distinct cases listed below.

(i) 3rd person conjunct DPs- Number agreement happens with the ConjP (and default gender value); examples (1-2).

Step 1: Agree: Predicate, Number (ConjP)- resolved
2:Copy value: Predicate, Number(ConjP)
3: Agree: Predicate, Gender (ConjP)- unresolved
4: Insert default gender value (masculine)

(ii) 3rd person conjunct DPs- CCA obtains for both number and gender features; examples (3-4).

Step 1: Agree: Predicate, Number (ConjP)- unresolved
2: Choose the linearly close conjunct
3. Agree and Copy the number value from this conjunct
4: Agree: Predicate, Gender (ConjP)- unresolved
5. Repeat steps 2 and 3 for gender value from the closest conjunct

(iii) Atleast one 1st/2nd person DP: Agreement takes place with the most marked person feature, followed by agreement with the conjP in number and gender features; examples (5-8).

Step 1: Agree: Predicate, Person (ConjP)- unresolved
2. Choose the conjunct with the most marked person feature
3. Agree and Copy the person value from this conjunct
4: Agree: Predicate, Number (ConjP)- resolved
5. Copy value: Predicate, Number (ConjP)
6: Agree: Predicate, Gender (ConjP)- resolved
7: Copy value: Predicate, Gender(ConjP)

Note that even if the number and gender values are unresolved at the ConjP in steps 4 and 6 above, no CCA can obtain. Default number and gender values are inserted

To sum up, I have presented hitherto undiscussed patterns of conjunct agreement with predicative adjectives in Punjabi, with focus on the role played by the person feature. CCA, which is generally obtainable with 3rd person conjunct DPs, fails to obtain altogether in the presence of a conjunct DP bearing a 1st/2nd person feature. Further, this work also lends support to split phi probing, such that person feature gets probes and gets valued before number feature, which precedes gender feature (in line with Bejar & Rezac 2003; Chomsky 2000; Taraldsen1995etc.).

References

Bhatia, A. (2011) Agreement in the Context of Coordination: Hindi as a Case Study. Doctoral Dissertation. University of Illinois.

Bejar,S. and Rezac, M. (2003) Person licensing and the derivation of PCC effects; In Romance Linguistics: Theory and Acquisition. 49–6. John Benjamins, Amsterdam

Benmamoun, E., A. Bhatia & M.Polinsky (2009) Closest Conjunct Agreement in Head Final Languages. Linguistic Variation Yearbook .67-88. Amsterdam/Philadelphia: John Benjamins.

Chomsky, N. (2000) Minimalist Enquiries. In Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik. Cambridge Mass: MIT Press.

Marušič, F., A. Nevins & A. Saksida (2007) Last-ConjunctAgreement in Slovenian. In Annual Workshop on [Formal] Approaches to [Slavic] Linguistics: The Toronto Meeting 2006, 210-227.

Ann Arbor: Michigan Slavic Publications.

Marušič, F, A. Nevins & W. Badecker (2010) The Grammars of Conjunction Agreement in Slovenian .Accessed at http://ling.auf.net/lingBuzz/001131.

Munn, A. (1999) First Conjunct Agreement: Against a Clausal Analysis. Linguistic Inquiry 30:4, 643-668.

Silverstein, M. (1976) Hierarchy of Features and Ergativity. In Grammatical Categories in Australian Languages, 112-171, Canberra: Australian Institutes of Aboriginal Studies.

Taraldsen, T. (1995) On agreement and nominative objects in Icelandic. In Studies in Comparative Germanic Syntax, 307–327. Dordrecht: Kluwer.

Some hidden features of Ergativity in Indo-Aryan languages: A case study of Urdu, Punjabi and Hindko

Zafeer Hussain Kiani
University of AJ&K,
Muzaffarabad
zafeerkiani@gmail.com

This paper aims to analyze split ergative behavior of three of the Indo-Aryan Languages i.e. Urdu, Hindko and Punjabi following the standard mechanism of case marking and verb agreement. There is a prevailing notion that Ergativity is subject to the transitivity of verb and perfectivity. The findings of this study support the claim that for ergative case a transitive verb should be in perfective aspect. However, the study explores another fact that a transitive verb along with perfective aspect can exhibit nominative case as well. The data from all three said languages help in establishing this fact. The condition which is necessary for ergative case to occur not always marks ergative case marker on the subject. Ergative case is also subject to the choice of light verb in a complex predicate.

In Indo-Aryan languages, according to Bhatt (2007), there are mostly two cases: the marked ergative case and unmarked nominative/accusative case. He reports that these languages exhibit aspect based split ergativity. We have ergative alignment only when the tense is perfective. However, some of these languages; like Kashmiri, also display person based ergativity. This section deals with a detailed study on the ergativity in the following three languages:

1 Urdu

Urdu is also a split ergative language (Butt; 1995, 2006) and the ergativity is conditioned to transitive verb and perfective aspect. She reports that the subject in Urdu is marked with ergative case when these two conditions are fulfilled. However there are certain instances in Urdu when these two conditions are fulfilled but still the subject exhibits nominative case. Consider the following examples:

1. farhan-ne sabak yad kar lia ha
   farhan-ERG lesson learn do take be
   "Farhan has learnt the lesson."

2. farhan sabak yad kar chuka ha
   farhan-NOM lesson learn do PERF be
   "Farhan has learnt the lesson."

3. momna-ne sabak yad kar lia ha
momna-ERG lesson learn do take be
"Momna has learnt the lesson."

4. momna sabak yad kar chuki ha
momna-NOM lesson learn do PERF be
"Momna has learnt the lesson."

The contrast in (1) and (2) shows that both the constructions convey the same meaning with two alternatives. Similarly, the nature of transitivity and perfectivity is also same in these two constructions but the subject in (1) is marked with the ergative case while in (2), it exhibits unmarked nominative case. The light verb lia-take agrees with the nominative sabaq-lesson in gender when the subjects were marked with ergative case in (1) and (3). When there was unmarked nominative case on subjects in (2) and (4). It shows that the choice of light verb in the complex predicate construction does affect the case marking on subject. It is interesting to note that this phenomenon happens only when there is a double verb construction. If the clause consists of a single verb with the conditions of ergativity be fulfilled, the subject will bear ergative case. Consider the following examples:

5. farhan-ne saim-ko kitaab bheji
farhan-ERG saim-ACC book se
"Farhan sent a book to Saim."

6. areeba-ne kapRe dhoye
Areeba-ERG clothes wash.PST
"Areeba washed the clothes."

7. hafsa-ne sabak yad kia
hafsa-ERG lesson learn do.PST
"Hafsa learnt the lesson."

The clauses in the examples (5-7) consist of a single verb and the transitive verb exhibits perfectivity, therefore the subject bears ergative case.

2 Punjabi

Punjabi shows the same behavior towards ergativity as Urdu does. The verb in Punjabi agrees with the subject if it bears nominative case (Akhtar, 2000) but when the transitive verb exhibits perfectivity, the subject is marked with the ergative case and thus the verb agrees with the other nominative in the construction. The following examples show this phenomenon:

8. aqsa-ne ansa-nu yaad kita
aqsa-ERG ansa-ACC think do.PST
"Aqsa thought about Ansa."

9. majid-ne rahat-da haal
majid-ERG rahat-GEN condition
puchhya
ask.PST
"Majid asked about the well being of Rahat."

It is noteworthy that this all is true for the clauses having single verb only. If a clause consists of double verb construction, the case on subject depends on the choice of light verb in the complex predicate construction. Consider the following examples:

10. basit-ne roti kha lei
basit-ERG meal eat take.PST
"Basit has taken his meal."

11. basit roti kha betha
basit-NOM meal eat sit.PST
"Basit has taken his meal."

12. sara-ne kəm mukaa lea
sara-ERG work finish take.PST
"Sara has finished the task."

13. sara kəm mukaa bethi
sara-NOM work finish sit.PST
"Sara has finished the task."

The contrast in (10) and (12) shows the same phenomenon that the subject in Punjabi also bears ergative case on the fulfillment of same conditions which are mandatory in Urdu .i.e. transitivity and perfectivity on part of the verb. On the other hand, the contrast in (11) and (13) shows that though the conditions are still fulfilled but the subject does not exhibit ergative case. The case marking on the subject is influenced by the selection of light verb in Punjabi too.

3 Hindko

Hindko language also belongs to the Indo-Aryan family of languages. Hindko is spoken by more than four million people (Crystal, 1997). It is spoken as native language by 2.4% of the total population of Pakistan (Addleton, 1986). It is the widely spoken language of district Muzaffarabad and Neelum valley, Azad Jammu & Kashmir.

The case system of Hindko is identical to those of case systems in Urdu and Punjabi. The subject bears ergative case when the clause exhibits transitive verb and perfective aspect. This happens only when the clause consists of single verb predicate. Following examples show the ergative case marking on subjects:

14. danish-sun ruti khadi
danish-ERG meal eat.PST
"Danish took his meal."

15. omer-sun eh shararat kiti
omer-ERG this mischief do.PST
"Omer did this mischief."

Like Urdu and Punjabi, when the Hindko clause consists of two verb predicate, the choice of case marking is conditioned to the choice of light verb in the complex predicate. Consider the following examples:

16. khuram-sun cha chhoR diti
khuram-ERG tea leave give.PST
"Khuram has abandoned taking tea."

17. anjum roti kha reya
anjum-NOM meal eat live.PST
"Anjum has taken his meal."

18. maira-sun kəm mukaa chhoRya
maira-ERG work finish leave.PST
"Maira has finished the task."

19. sadaf kəm mukaa reyi
sadaf-NOM work finish live.PST
"Sadaf has finished the task."

The examples (16-19) exhibit transitivity and perfectivity but the subject in (17) and (19) bear nominative case while there is ergative case on subject in (16) and (18). This shows that ergative case marking in Hindko is also conditioned to the light verb in complex predicate.

4 Conclusion

The main focus of this study was on ergative case marking in Urdu, Punjabi and Hindko. The case marking in the said languages was discussed in previous research. This study particularly observed the conditions which are necessary for the ergative case marking. For this purpose, various examples from these languages were examined following the pattern of case marking and verb agreement. The results show that though transitivity and perfectivity are the necessary conditions in a clause to exhibit ergative case marking but it is not necessary for any clause to exhibit ergative case on subject having these two conditions fulfilled. The subject may bear nominative case with the same conditions which are necessary for ergative case. The ergative case on subject is also conditioned to the light verb in a complex predicate construction. However, the results also show that if a clause with a simple predicate fulfills two said conditions for ergativity, the subject will bear ergative case. Only the subject of a complex predicate construction is influenced by the light verb in assigning the case.

References

Addleton, J. S. 1986. The importance of regional languages in Pakistan. Al-Mushir XXVIII 2, 58-80.

Akhtar, R. N. 2000. Aspectual comple predictes in Punjai (Unpublished Doctoral Dissertation). University of Essex, Colchester.

Bhatt, R. 2007. Ergativity in the Modern Indo-Aryan Languages MIT Ergavitivity Seminar.

Butt, M. 1995. The Structure of Complex Predicates in Urdu. CSLI Publications,Stanford, California.

Butt, M. 2006. The Dative-Ergative Connection. In O. Bonami & P. Cabredo Hofherr (Eds.) Empirical Issues in Syntax and Semantics, 6. (pp. 69–92)

Crystal, D. 1997. The Cambridge Encyclopaedia of Language. Cambridge: CUP.

Aspectual Classes and the semantics of the Hindi-Urdu verb

Rainer Kimmig
University of Tübingen
rainer.kimmig@uni-tuebingen.de

1 Introduction

Even a superficial glance at the verb entries in a dictionary of Hindi or Urdu (subsequently HU) will show that many lexems are much 'broader' semantically than the corresponding ones in European languages. This observation is of particular relevance when it comes to mutually exclusive aspectual realisations of a given verb. Thus, one finds, for instance, for *honā* both 'be' and 'become', for *jānnā* both 'know' and 'learn', and for *ḍhũ̃ṛhnā* both 'search' and 'find' (examples from McGregor 1993).

The present paper will argue that, in a given context, the restriction to just one specific meaning is conditioned by an interplay of both grammatical constraints and the semantic properties of the particular verb.

2 Vendler and the HU verb

A rough, but nonetheless useful conceptual tool for the analysis of the partial meanings of verbs like *honā*, *jānā* and *ḍhũ̃ṛhnā* is provided by the four classes set up by Vendler (Vendler 1967):

1. states: (e.g. *know*)
2. activities (e.g. *run, draw*)
3. accomplishments (e.g. *build, draw a circle*)
4. achievements (e.g. *find, reach the summit*)

Vendler's distinctions are strictly based on 'time schemata', i.e. temporal notions as duration and boundedness (for details, see below). For a clarification of the intricacies of the semantics of HU verbs, a fine-tuning is necessary in at least two points:

1. As Verkuyl pointed out, an analysis of aspectual properties, particulary in the case of activities and accomplishments, should rather consider sentences than just verbs (Verkuyl 2012: 575; but see already Vendler's distinction between *draw* and *draw a circle*).

2. The concepts of activity and accomplishment are too broad and therefore in need of further specification. Here, I rely on the work of Maslov (1984), Bondarko (1995) and Zaliznyak & Šmelev (1997), particularly on their discussion of the concept of telicity (or 'terminativity'), which is indispensable both for the analysis of Russian aspect (*vid*) and the HU compound verb. *States and activities are atelic, accomplishments and achievements are telic* (Bondarko 1995: 33–48; Zaliznyak, Šmelev 1997: 45–58).

The following sections will show that many HU verbs are ambivalent with respect to Vendler's classes, selecting one class in a particular grammatical environment and another one in another.

3 Verbs denoting either states or accomplishments

A common type of states is expressed in HU by verbs in the simple present or in tense forms related to it: *vah jāntā hai* 'he knows', *vah jāntā thā* 'he knew'.

Verbs of this type show a clear-cut contrast between the tense forms denoting states and others denoting a change of state, i.e. accomplishments (*vah jān rahā hai* 'he is getting to know, he is learning'):

- *state* – simple present, habitual past
- *accomplishment* – progressive tenses, simple past, compound verbs

Note: the accomplishment reading is compatible both with the progressive and the perfective (compound verb).

The following examples may illustrate this.

Contrast between state (simple present) and accomplishment (progressive present):

- *maĩ film-mekar Anurāg ko to jāntī hũ̃. lekin bataur pati unhẽ dhīre-dhīre jān rahī hũ̃.*
  (BBC Hindi)
  I do know the film-maker Anurag. But I am getting to know him only slowly as a husband.

Change of state, accomplishment (simple past):

- *āj pahlī bār jānā ki parīkṣa kya hotī hai.*
  (Navbharat Times)
  Today, he realized for the first time what an exam really was.

Change of state, accomplishment (compound verb, e.g., *jān jānā* 'come to know', in all tense forms possible for compound verbs):

- Simple present:
  pahle ham ek savāl karte haĩ aur bād mẽ iske uttar ko jān jāte haĩ. (Ādhārbhut Brahmāṇḍa)
  First, we ask a question and later we find the answer.
- Present perfect:
  maĩ tumhāre ēṭīem kā pāswarḍ jān gayā hũ̃.
  (Hindi Jokes)
  I have found out your ATM pin.
- Conditional (contrasting with the simple present and the conditional of the simple verb):
  jāntā hũ̃. aur na bhī jāntā hotā to jān jātā.
  (Pallavī. Hindī Pāṭhya pustak 8)
  I know. And even if I did not know, I would come to know.

4 Verbs denoting either activities or their aspectual counterparts

Activities, according to Vendler, are processes or actions that go on for some time without approaching or reaching a limit (*sleep, live, run, draw, read*).

HU Verbs denoting activities in some tenses, may denote processes or actions restricted by a boundary in other tenses or in compound verbs. I will discuss two distinct cases.

1. Activities vs. inchoatives: *calnā* 'move' and 'start moving' (Zaliznyak, Šmelev 1997: 51, 89–94)

- *activity* – progressive tenses
- *achievement, inchoative* – simple past, compound verb
- both readings are possible in the simple present, depending on context

Activity, ongoing action:

- *uske pīche pajero gāṛī cal rahī thī.*
  (Dainik Bhaskar)
  He was followed by a Pajero car ('a Pajero car was driving behind him').

Achievement, onset of action:

- simple past, compound verb:
  gāṛī calī
  gāṛī cal dī
  The train left / The car startet.

2. Activities or inchoative achievements vs. deliminatives: *sonā, so jānā* 'sleep, fall asleep' vs. *so lenā* 'sleep for a while, have a nap' (Zaliznyak, Šmelev 1997: 94)

- *activity* and *achievement, inchoative* – simple verb (*sonā*), compound verb (*so jānā*); the contrast is the same as with *calnā, cal denā*
- *accomplishment, deliminative aktionsart* – simple past, compound verb with V2 *lenā*

The aspectual contrast in this subgroup has an exact correspondance in the contrast between Russian imperfective activities and deliminative perfectives formed with the prefix *no*. Compare:

sonā спать^{i} 'sleep'
– *so lenā* поспать^{p} 'take a nap'[11]

ghūmnā гулять^{i} 'walk, stroll'
– *ghūm lenā* погулять^{p} 'have a (short) walk'

jīnā жить^{i} *'live'*
— *jī lenā* пожить^{p} 'live (for a short while)'

[11] Following Forsyth 1970, I use the superscript letters ^{i} (imperfective) and ^{p} (perfective) to indicate the aspect of Russian verbs.

- *roz savere maĩ thoṛā-sā atīt mẽ jī letā hū̃.*
  (Ajneya)
  Every morning I live a little in the past.

5 Verbs denoting accomplishments

Accomplishments, according to Vendler, are telic processes or actions that take some time to reach their limit (*get exhausted, grow old, cure, build*). In other words, they denote a 'gradual approach to a result' (Zaliznyak, Šmelev 1997: 52). Activities turn into accomplishments when limited by a boundary, e.g., an object: *read a book, draw a circle*.

A particularly interesting subgroup of accomplishments expresses the contrast between attempt and success (Zaliznyak, Šmelev 1997: 20, 49, 52):

- *hamne logõ ko bahut samjhāyā, lekin kyā karẽ, logõ ko samjhānā bahut muśkil hai.*
  (Dr. Harsh Vardhan)
  We tried hard to convince the people, but what can we do, it's so difficult to convince the people.

With verbs of this type, three cases have to be distinguished in the non-progressive past tenses:

- *attempt, mostly, but not necessarily successful* – simple verb
- *'cancelled' success, conative* – simple verb
- *success* – compound verb

Compare the use of *ḍhū̃ṛhnā* 'search' and 'find (after a long search)':

- *simple verb, successful attempt:*
  Amrīka ne aise ḍhū̃ṛhā Bin Laden ko.
  (BBC Hindi)
  How America found Bin Laden (after an almost ten year long search!)
- *simple verb, 'cancelled' success:*
  aisā ghar unhõne bahut ḍhū̃ṛhā, par na milā.
  (*Premchand*)
  He had long searched for such a home, but he had not found one.
- *Compound verb, success:*
  is din vaijñānikõ ne ghoṣṇā kī ki unhõne 'higs boson' ke nām ke ek aise sūkṣma kaṇ ko dhū̃ṛh nikālā hai jiskī khoj mẽ pichle kaī daśak se vaijñānik lage hue the. (Mazdūr Bigul)
  On this day, scientists have announced that they had found a particle called 'Higgs boson', in the search of which scientists had been engaged for the last several decades.

The clear-cut distinction between cancelled success and success in and has an exact analogy in Russian. Like the compound verb in HU, only the perfective aspect in Russian implies that a telic action reached its result. On the other hand, in Russian would prefer the perfective. Compare:

- *Так Американцы нашли[P] ядерное оружие у Саддама?* (otvet.mail.ru)
  So the Americans did find nuclear weapons with Saddam?

6 Achievements

Perhaps the most problematic among Vendler's classes are achievements. Achievements are limited to 'single moments of time (strictly speaking)' (*reach the top, win a race*). Thus, in English, they do not occur in the progressive tenses. Most HU verbs which, in their core meaning, do qualify for such a definition (e.g. *pahũncnā* 'arrive at, reach') have broader usages, too, which may be classified as accomplishments.

However, with respect to *pahũncnā* in its core meaning of 'arrive at' it can be observed that:

— The progressive does denote an actual present, but a scheduled future:

- *maĩne apne sampādak mitra ko tār diyā thā ki amuk havāī jahāz se karācī pahũc rahā hũ̃, aur do din tak, bībī baccõ ke sāth tumhārī roṭī toṛũ̃gā.* (Bhisham Sahni)
  I had sent a cable to my editor friend that I was arriving at Karachi on flight so and so and that I would share his bread for two or three days together with my family.

— In a situation of coincidence, the simple present is used rather than the progressive tenses:

- *pleṭfarm pahũcte hī the ki gāṛī cal dī* (Naidunia)
  They just reached the platform when the train left.

7 Conclusion

The rough outline given above shows that a description of the HU verb in terms of a refined Vendlerian classification is relevant both in grammatical analysis and lexicography.

References

Bondarko, A.V. 1995. *Die Semantik des Verbalaspekts im Russischen. Семантика глагольного вида в русском языке.* Frankfurt et al: Peter Lang.

Forsyth, J. 1970. *A Grammar of Aspect. Usage and Meaning in the Russian Verb.* Cambridge: Cambridge University Press.

Maslov, Y. 1984 = Маслов, Ю. *Очерки по аспектологии.* Ленинград: Издательство Ленинградского Университета.

Hook, P.E. 1973. *The Compound Verb in Hindi.* Michigan: The University of Michigan, Center for South and Southeast Asian Studies.

McGregor, R.S. 1993. *The Oxford Hindi-English Dictionary.* Oxford Delhi: Oxford University Press.

Nespital, H. 1997. *Dictionary of Hindi Verbs.* Allahabad: Lokbharati Prakashan 1997.

Vendler, Z. 1967. "Verbs and Times." In Z. Vendler *Linguistics in Philosophy.* Ithaca and London: Cornell University Press.

Verkuyl, H.J. 2012. "Compositionality". In Binnick (ed.). *The Oxford Handbook of Tense and Aspect.* Oxford et al.: Oxford University Press.

Zaliznyak, A.A., Šmelev, A.D. 1997 = Зализняк, А.А., Шмелев, А.Д. *Лекции по русской аспектологии.* München: Otto Sagner.

Causative Alternations in Tenyidie

Kikrokhol Kraho
University of Hyderabad
kikrokhol@gmail.com

1 Introduction

Causative constructions continue to attract linguists because, as Comrie (1989) observed, their 'study involves the interaction of various components of the overall linguistic description, including semantics, syntax, and morphology'. In this paper, I describe in brief, the different causative alternations found in Tenyidie.[12]

2 Lexical causatives

As in most languages, Tenyidie also has a number of verb pairs whose members describe similar situations but differ in that one of them has one more theta role than the other. Some examples are given below.

cü 'to eat' *vachü* 'to feed'
kre 'to drink' *die* 'to feed with liquid'
co 'be awake' *kesu* 'to wake'
tu 'burn' *the* 'to burn'

There is no formal relation between the members of the pairs given above, and it cannot be said that the causative verbs are derived from the non-causative verbs or vice versa. The pairs differ in an unpredictable manner. As Cole (1983) noted for similar data in Kannada, sentences with causative meaning are most felicitous when the object of the verb is, for instance, an animal or a child or an invalid, i.e., 'someone to whom volition is not imputed'.

3 Transitive unaccusative alternations

Levin and Rappaport Hovav (1995) observe that verbs that are found in causative alternations are the verbs depicting change of state, which typically describe changes in physical shape or appearance. Based on Jespersen's (1927) characterization, they group such verbs into two different sub-classes: verbs like *bounce, move, roll, rotate, spin* are called "motion" class verbs and those like *bake, blacken, break, cook, cool, dry, freeze,* etc., are called "change" class verbs. In Tenyidie, these two different verb classes display a systematic morphological distinction in transitive-unaccusative alternations.

All unaccusatives member of the 'move' class – *rüve* 'rotate', *rütou* 'roll', *rülhou* 'bounce' etc. – are disyllabic, with an empty morph *rü* in the word initial position. The transitive members of the class are also disyllabic but differently marked. They have *ke* in place of *rü* as shown in (2) below.

1. A hu rünyie ba te
   1SG tooth shake sit PFT
   'My tooth is shaking.'
2. Puo a hu kenyie shü
   2SG 1SG tooth shake PFV
   'He shook my tooth.'

The transitive alternants obligatorily select an animate subject. Therefore (3) is ungrammatical. However, there are also situations where natural forces like wind or other inanimate objects and phenomena can be assumed to cause other events. When such situations are described, another verb – which describes the causing event – is added as in (4). The two verbs together behave as a single predicate.

3. * Teikhrie-e ki u kenyie shü
   win-EMP house DET shake PFV
   'The wind shook the house.'
4. Teikhrie-e ki u he kenyie shü
   win-EMP house DET blow shake PFV
   'The wind shook the house.'

The unaccusative verbs of the "change" class appear in different syllabic structures; some are monosyllabic and some are disyllabic. All the monosyllabic verbs begin with voiceless stops, and are related to the verb 'break'.[13] The disyllabic verbs begin with *rü* and they describe changes in physical shape. The transitive members of the alternation are not regularly marked as they are in the 'move' class but the marking is not random either. The monosyllabic verbs are aspirated and appended to another verb which describes the causing event, as in (6).

5. Si u tse te
   stick DET break PFT
   'The stick broke.'
6. A si u ki-ts^{h}e shü
   1SG stick DET pull-break PFV
   'I broke the stick (by pulling it).'

The aspirated forms appear as bound morphemes and they do not occur as independent verbs with the same meaning elsewhere. So, (7) is ungrammatical.

7. *A si u ts^{h}e shü
   1SG stick DET break PFV
   'I broke the stick.'

The transitive alternants of the disyllabic verbs are encoded in the following manner: in place of *rü* another activity verb appears, and if the second syllable begins with a voiceless stop, it is aspirated as in (9); if not, it remains unchanged as in (11).

8. Ketsie u rüko te
   stone DET crack PFT
   'The stone cracked.'
9. Vizo ketsie u vü-kho shü
   vizo stone DET hit-crack PFV
   'Vizo cracked the stone (by hitting it)
10. Gali u rüde te
    pot DET deform PFT
    'The pot deformed.'
11. Niu gali u tsa-de shü
    niu pot DET trample-deform PFV
    'Niu deformed the pot (by trampling upon it).'

There are a few verb pairs like *rüzou* 'fallout from the sheath' and *zou* 'to remove from the sheath' where the second syllable of the unaccusative verb appears by itself without any change in the transitive counterpart.

[12] Tenyidie belongs to the Tibeto-Burman language family. It is spoken in India, particularly in and around the state of Nagaland.

[13] Tenyidie has four different verbs for 'break', and usage depends on the kind of object involved in the action of breaking.

4 Causative alternation

Causative verbs are derived by prefixing the causative morpheme to a non-causative verb. Like the other verbal pairs, the derived causatives and its corresponding non-derived form describe similar situations, but they differ in that one verb describes a situation which can happen only in the presence of a causer and a causee. The other describes a situation that does not require a causer. The derived verbs are uniformly marked with the prefix *pe-*. The prefix occurs with intransitive as well as transitive verbs, however like the morpheme *aa-* in Hindi, as in *chaakh-na* "to taste" and *chakh-aa*-na "to cause to taste", only a handful of transitive verbs, the so-called *ingestive* verbs, take this morpheme (Ramchand 2008).

The prefix can occur with verbs that have non-derived causative alternants. Thus, the derived causative verb form exists side by side with the non-derived causative as shown below.

pe-co 'cause to wake' *kesu* 'to wake'
pe-tu 'cause to burn' *the* 'to burn'
pe-sie 'cause to die' *dukhri* 'to kill'

Both verbs in each pair describe a causative situation but with a distinction in interpretation. The derived forms roughly mean *cause to V-intransitive* (Levin and Rappaport Hovav 2005). The non-derived members of the pairs mean something like *α executes the action on β* (Horvath and Siloni 2010). (12) means the noise caused the speaker to wake up whereas (13) means Viu intentionally woke the speaker up. The verb *kesu* 'to wake' takes a human agent to directly execute the action, therefore (14) is ungrammatical.

12. Mhapfe u a pe-co shü
noise DET 1SG CAU-wake PFV
'The noise woke me up.'
13. Viu a kesu shü
viu 1SG wake PFV
'Viu woke me up.'
14. *Mhapfe u a kesu shü
noise DET 1SG wake PFV
'The noise woke me up.'

There is a group of unaccusative verbs that never take the prefix. These are verbs that have another corresponding derived causative verb form. So, all of the following are ill-formed:

**pe-rünyie* 'cause to shake'
**pe-rüve* 'cause to spin or rotate'
**pe-tse* 'cause to break'
**pe-rüko* 'cause to crack'
**pe-rüwi* 'cause to bend'

The restriction here is reminiscent of one seen in Marathi where the causative suffix *-aw* – which has a moderately high degree of productivity – gets attached only to those intransitive verbs that do not have other causative forms. In Marathi **ughaD-aw-Ne* 'to open ', **khaa-aw-Ne* 'to feed', **mar-aw-Ne* 'to kill', are not possible because there are other derived verb forms to express these meanings (Shibatani and Prashant 2002).

When the prefix *pe-* occurs with unergative verbs like *ta* 'run', *kra* 'cry', *pro* 'fly' etc., the meaning comes close to *α causes β to do the action*, as in (15).

15. Lima pera u pe-pro wa te
lima bird DET CAU-fly PVT PFT
'Lima caused the bird to fly away.'

Among the transitive verbs only a few appear with the causativizing morpheme. They are *si* 'know' *ngu* 'see', and *chü* 'hear'. So we have the following pairs.

si 'know' *pe-si* 'cause to know/to inform'
ngu 'see' *pe-ngu* 'cause to see/to show'
chü 'hear' *pe-chü* 'cause to hear'

These verbs appear to form a natural class like the *ingestive* verbs in Hindi and it can be said that *pe-* is like the causativiser *aa-* in Hindi.

5 Conclusion

This is an attempt to provide a brief account of the surface facts pertaining to the different kinds of causative alternations in Tenyidie, rather than trying to account for these facts within the framework of any particular linguistic theory. The examples show that Tenyidie is similar to other languages in many respects. It conforms to the different observations made by different writers on the phenomenon of causative alternations in other languages. One notable fact about Tenyidie is that the 'move' class and the 'change' class of causative alternations are differently marked. This provides evidence for the claim that verbs that participate in causative alternations fall into two distinct classes.

Reference

Cole, Peter. 1983. "The Grammatical Role the Causee in University Grammar". *Internal Journal of American Linguistics* 49:155-1333.

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. Chicago: University of Chicago Press.

Horvath, Julia, and Tal Siloni. 2010. "Lexical versus Syntax: Evidence from Morphological Causatives". In Malka Rappaport Hovav, Edit Doron, and Ivy Sichel (eds.) *Lexical Semantics, Syntax and Event Structure*. Oxford: Oxford University Press.

Jesperson, Otto. 1927. *A modern English Grammar on Historical Principles. Part 3: Syntax, Second Volume.* Heidelberg: Carl Winter.

Levin, B. and M. Rappaport Hovav. 1995. *Unaccusativity*: *At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.

Ramchand, Gillian Catriona. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. London: Cambridge University Press.

Rappaport Hovav, Malka and Beth Levin. 2005. "Change of State Verbs: Implications for Theories of Argument Projection". In N. Erteschik-Shir and T. Rapoport (eds.) *The Syntax of Aspect*. Oxford University Press.

Shibatani, Masayoshi and Prashant Pardeshi. 2002. "The causative continuum". In Masayoshi Shibatani (ed.) *The Grammar of Causation and Interpersonal manipulation.* Philadelphia: John Benjamins.

Influence of Malayalam on the verbal system of the Indo-Portuguese creoles of the Malabar

Ana Krajinovic Rodrigues
University of Lisbon, Portugal
anakrajinovic@hotmail.com

1 Introduction

As a consequence of Portuguese colonialism and economic role in Asia in the 16th century, language contact between Portuguese and local languages resulted in the creation of many creole languages along the Asian coast. Despite having once been especially numerous in the Indian subcontinent, Indo-Portuguese creoles entered a phase of obsolescence and nowadays are given practically as extinct. However, recent fieldwork in India has provided some linguistic data from various creoles, including the Indo-Portuguese creoles of the Malabar coast.

These languages were once spoken in a wide stretch of the Malabar coast, situated on the southwestern coast of India. Today, however, they are limited to a few last speakers living in Cannanore [Kannur] but, as a result of fieldwork carried out from 2006 onwards, we have recordings of four speakers, one of whom appears to have been the last speaker of Indo-Portuguese in Cochin [Kochi].

Early language contact in the Malabar in the 16th century, which led to the creation of Indo-Portuguese creoles, suggests the involvement of Portuguese as a superstrate language and of Malayalam, a language of the Dravidian family, as a substrate language. The Indo-Portuguese creoles of the Dravidian space have been known for their convergence towards the Dravidian substrate and adstrate (Smith 1977), which puts them in a rather unique situation typologically speaking. It is therefore important to determine the nature and extent of the influence of Malayalam on the verbal system in the Indo-Portuguese of Malabar, which will bring us closer to understanding the tendencies of the morphological and syntactic development of Indo-Portuguese creoles in a Dravidian environment. Moreover, in a broader context, this should give us a clearer idea of how powerful the influence of substrate and adstrate languages is, particularly when applied to languages with relatively rich morphology, as the ones of the Dravidian family. Indeed, it has already been shown by Cardoso (2014) that there is a high degree of Malayalam influence on Indo-Portuguese syntax and nominal structures. Our goal is to do a comparative analysis of the creole, Portuguese and Malayalam verbal systems in order to examine the actual extent of the influence of Malayalam on the Indo-Portuguese verb-related grammar.

2 Comparative analysis

As part of the ongoing study of the Indo-Portuguese grammar in relationship with Portuguese and Malayalam, we compare the basic Malayalam, Portuguese and Indo-Portuguese strategies to express tense, aspect, modality and some specific verbal functions that cannot be immediately attributed to a Portuguese source. Even when the etymological source of a certain marker is Portuguese, its functional distribution often follows a different model. For instance, we can easily deduce a Malayalam origin of the Indo-Portuguese pre-verbal marker *lɔ* — a junction of future tense and irrealis mood — by noting that the future tense morphology in Malayalam can often be linked to the expression of various irrealis modalities. As we can see in example (1), *lɔ* is used to denote future tense in Indo-Portuguese.

(1) *olotrə uŋa kaza lɔ fazæ.*
they one house FUT make
'They will build the house.'
(Cannanore: oral corpus)

Based solely on example (1), the future tense could be interpreted as displaying either Portuguese or Malayalam influence. However, the marker *lɔ* can also appear in semantic domains of irrealis modality, such as the notion of possibility, as in example (2).

(2) *yo lɔ vay.*
I FUT-IRR go
'I will go / I may go.'
(Cannanore: oral corpus)

In Malayalam there is an overlap between future tense and modal morphology. Namely, by using different future tense suffixes, the speaker can indicate the degree of certainty with which he makes an assertion (Asher and Kumari 1997:311). In example (3), for instance, we find the modal suffix *-aam* expressing a strong commitment to a future course (Asher and Kumari 1997), like the Indo-Portuguese *lɔ*-construction in example (1).

(3) *ɲaan budhanaaʐca varaam*
I Wednesday come-FUT-MOD
'I will come on Wednesday.'
(Asher and Kumari 1997:290)

At the same time, a lower degree of certainty (or a possibility) may be expressed in Malayalam by using the future suffix *-(eekk)um*, as is clear from example (4).

(4) *aɻutta kollam ɲaan heɖmaasttar aayeekkum*
next year I headmaster be-FUT-MOD
'Next year I may become a headmaster.'
(Asher and Kumari 1997:102)

By comparing examples (2) and (4), we conclude that the Indo-Portuguese use of future morphology to express uncertainty or possibility may have derived from the said "scale of uncertainty" in Malayalam.

Another point of comparison of the usage of future tense in Indo-Portuguese and Malayalam is that of conditional sentences. In Indo-Portuguese, in both factual and counterfactual conditional sentences, the future tense with the marker *lɔ* seems to be chosen in the main clause. This is particularly striking in the case of counterfactual conditional sentences, as in example (5), where a *lɔ*-construction appears in the main clause, referring to an irrealis event.

(5) *akəl ja da sə paymi ũga*
DEM PAST give COND 1s.OBL one
mil rupi lɔ da.
thousand rupee IRR give
'If [I] delivered [i.e., a skin], [they] would give me one thousand rupees.'
(Cardoso 2014:313)

This Indo-Portuguese use of future tense in conditional sentences can be compared to the Malayalam use of future morphology in both factual and counterfactual conditional sentences. In the latter case the verb in the main clause, besides containing future tense morphemes, also has past tense morphology, as we can see in example (6).

(6) *avaɭ nallavaŋŋam paɻhiccirunneŋkil*
she well study-PERF$_1$-PAST-COND
jayikkumaayirunnu
pass-FUT-PAST
'If she had studied well, she would have passed.'
(Asher and Kumari 1997:89)

Although Malayalam exhibits a more specific tense morphology than the creole in conditional sentences, the use of the future morphemes in the main clause can nevertheless be recognised as a common characteristic.

Based on this data, it is clear that the Indo-Portuguese connection of future and irrealis must proceed from a Malayalam source rather than a Portuguese one, since Portuguese future tense morphology does not express any modal semantics in this way.

There are in fact many other Indo-Portuguese verbal functions that seem to be derived from Malayalam strategies rather than Portuguese, such as some forms of expressing verbal aspect or syncretism of copula, possessive and existential elements. Our analysis will encompass all such cases which can be identified in the creole and Malayalam data.

References

Asher, R.E. and T.C. Kumari. 1997. *Malayalam* [Descriptive Grammars Series]. London/New York: Routledge.

Cardoso, Hugo. 2014. *The case of addressees in Dravido-Portuguese. O caso dos interlocutores em Drávido-Português*. PAPIA, São Paulo, 24(2), p. 307-342.

Smith, Ian R. 1977. *Sri Lanka Creole Portuguese Phonology.* Doctoral dissertation, Cornell University

Synchronic and diachronic trends in phi-features of South Asian languages

Benjamin K. Macaulay
CUNY Graduate Center
bmacaulay@gc.cuny.edu

1 Introduction

South Asia has long been known as a linguistic area, where features spread geographically under contact, and are not necessarily the result of genetic inheritance.

One example is the retroflex consonant series, which is found in nearly all Indo-Aryan languages but is ultimately the result of contact with Dravidian languages and not an inherited feature from Proto-Indo-European.

This paper examines commonalities in South Asia as a linguistic area, not in phonology but in morphology. Evidence from South Asian languages has been used to inform the synchronic organization of phi-features (Macaulay, to appear), and this information can then be used as a window to how these synchronic structures arise and develop.

2 Phi-feature formality

Contrasting levels of formality is a common phenomenon in the world's languages. For example, French uses plural pronouns/verb markings as a formal singular in the 2nd person (called the "T-V Distinction after the pronouns *tu/vous*).

Languages vary in how this formality is expressed: plural-as-formal as in French, and 3rd person as formal 2nd person (Italian) are common. Because these strategies involve features like number/person, traditional accounts of phi-features treat formality as a pragmatic selection of morphemes with these other features, rather than a discrete feature (Corbett 2006, 2012, Harley & Ritter 2002b).

Many languages, however, cannot account for formality in their agreement relations purely through other features. Indo-Aryan and Dravidian languages in particular have many morphemes that are contrastive only for level of formality.

Bangla is a very clear example, where verb markings are specified for person and level of formality, but not number (or gender, etc.):

		formality		
		L	M	H
person	1	-i		
	2	-iʃ	-o	-en
	3	-e		

Table1: Bangla present tense verb marking paradigm

In Table 1, H, M, and L are used to refer to high, middle and low levels of formality respectively. Any of these markings can be used with a singular or plural referent (with an associative marker on the pronoun). A formality distinction is present in 3rd person in addition to 2nd. This means that Bangla's formality distinction cannot be derived from person/number alone. Levels of formality as some sort of gender/class feature are also not possible, for reasons to be discussed in Section 3.

3 Synchronic organization of phi-features

Because languages like Bangla have contrastive levels of formality that cannot be accounted for by means of other phi-features, formality must play a role in the synchronic organization of phi-features.

This paper assumes a morphosyntactic feature geometry, as proposed by Harley & Ritter (2002a, 2002b). Under this approach, phi-features are governed by multiple features under a REFERRING-EXPRESSION node. For example, person features have the following organization:

$$[_{\text{R-E}}\ [_{\text{PARTICIPANT}}\ \text{AUTHOR ADDRESSEE}]\ \dots\ ]$$

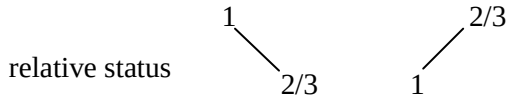
Here, the PARTICIPANT feature specifies that the morpheme refers to 1st/2nd person. Under this node, an AUTHOR feature specifies 1st person or ADDRESSEE for 2nd person. This derives typological findings such as evidence of 3rd person acting as a lack of features rather than its own feature (Nevins 2007). The ellipsis is where other nodes under R-E are found, namely INDIVIDUATION (number/gender) and CASE.

Macaulay (to appear) uses typological data to inform a representation of formality as a single binary feature. Namely, languages do not seem to exhibit both a formality distinction in 1st person and a three-level distinction (in any person). Macaulay derives this as a semantic overlap: if 1M/H and 2L/M/H are present in a language, does an utterance with 1M and 2L convey the same social dynamic as 1H with 2M? Can these contrast with an utterance with 1H/2L? This is a question the world's languages don't seem to ask. Maximally, a representation must account for *either* 1H or 2/3L, as shown in the largest paradigms in Macaulay (to appear), Lyélé (Niger-Congo, Burkina Faso) and Nepali (Indo-Aryan):

	Lyélé	Nepali
1st person – second level	✓	
2nd person – second level	✓	✓
2nd person – third level		✓
3rd person – second level	✓	✓
3rd person – third level		✓

Table2: Maximal formality paradigms

In order to account for either maximal paradigm but exclude their union, Macaulay uses a deictic approach to formality features that calculates the difference in social status between 1st and 2nd/3rd person:



Lyélé reflex	1H	2/3H
Nepali reflex	2/3L	H
	[+STATUS]	[-STATUS]

Table3: Deictic approach to formality

Under Macaulay's analysis, values of [+STATUS] denote situations where either the speaker's high status or the 2nd/3rd person's low status is stressed, and vice versa for [-STATUS]. The former situations surface in Lyélé as 1H but 2/3L in Nepali, while the latter situations show up as 2/3H in both.

Of note is that such an approach derives certain patterns of syncretism, whereby [+STATUS] is represented by a single morpheme in Lyélé and [-STATUS] by a single morpheme in many Indo-Aryan languages, for example the Bangla marking *-en* shown in Table 1 (or Nepali "H" in Table 3). Interestingly, as languages like Bangla are *pro*-drop, phrases with this marking may be ambiguous for 2nd/3rd person, but not level of formality.

Macaulay (to appear) incorporates [±STATUS] into Harley & Ritter's geometry as a sister node to PARTICIPANT, requiring a new node DYNAMIC to dominate person/formality features to the exclusion of other features:

$$[_{\text{R-E}}\ [_{\text{DYNAMIC}}\ [_{\text{PART}}\ \text{AUTH ADDR}]\ [\pm\text{STATUS}]]\ \ldots]$$

This analysis has implications for current theories of person features, notably that the new node DYNAMIC is now available for 3rd person, contradicting Nevins 2007. Also, all features that make use of deixis pattern as a group, while innate features like number and gender group together elsewhere. Additionally, a deictic analysis prevents interpretation of formality as an (innate) class feature.

The analysis predicts that languages that do not mark verbs for person cannot have agreement for formality. This is shown in Japanese, where verbs are marked for levels of formality but no mismatch is created with pronouns of different levels (Macaulay, to appear).

4 The domain of formality features

Macaulay (to appear) found evidence of formality features outside of verb markings, notably in honorific suffixes such as Hindi *ji*. A phrase with a subject suffixed with *ji* and a verb marked with M/L is ungrammatical. The reverse, failing to suffix *ji* on a subject agreeing with an H-level verb, was dispreferred but not judged fully ungrammatical by all informants. Preliminary data suggests a similar situation for the Tamil honorific *thiru*. This suggests that these morphemes are specified for featural content governing formality.

This is again in contrast to Japanese where honorifics like *sama* can freely combine with verbs of either register (informants found 'mismatches' rude or awkward at worst).

5 Diachronic paths of Indo-Aryan features

Formality features are redundant in many languages like French but essential to contrasts in languages like Bangla. The other Indo-Aryan languages fall on many points on the spectrum in between the two. These stages where formality is only partially redundant as a feature give insight into how paradigms like Bangla developed.

Hindi has verb markings cognate with the Bangla paradigm in Table 1. Bangla 2/3H *-en* corresponds to Hindi *-ẽ*, but Hindi uses the marking for informal 3pl as well as formal singular/plural of 2/3. Additionally, Hindi 2M *-o* (cf. Bangla 2M *-o*) is also a plural for 2L.

It seems that Indo-Aryan languages like Hindi started with a simple person/number system without a formality distinction, before licensing plural-as-formal and a third level using a 3rd person lexical substitution (German once had three levels, where plural-as-formal was M and 3rd person substitution was H, but the two have collapsed to 3pl=2M/H).

This created a redundancy in the paradigm, allowing some Indo-Aryan languages to begin to neutralize their number contrast, as Bangla has done fully. For example, Nepali has a very limited number contrast in verb markings: there is a separate 2/3L in the singular, and the only other contrast is between 1sg/1pl. Some speakers of Nepali do not distinguish between 1sg/pl, using 1pl *hami* and its verb marking *-aũ* for both. Informal sources suggest that such speakers use 1pl in all situations, not as a contrastive formality distinction. This seems to suggest that Nepali is on the path to collapsing its number distinction the way Bangla has.

6 Conclusion

Unrelated languages in South Asia such as Indo-Aryan and Dravidian languages share not only phonological but morphological features. Most striking are the large paradigms of formality that enter agreement relations with verbs and also other morphemes such as honorifics.

While these paradigms reflect trends cross-linguistic trends more generally, the large amount of variation in the expression of formality among Indo-Aryan languages allows us to see how these paradigms arise and evolve, as well as what effects the presence of phi-feature formality has on other features such as number.

Additional research will illuminate the role of phi-feature formality in the context of language contact and incorporate non-phonological linguistic features into work that models the process by which features spread areally (Blevins, to appear).

References

Blevins, J. to appear. "Areal Sound Patterns: From Perceptual Magnets to Stone Soup." In R. Hickey (ed.) *The Cambridge Handbook of Areal Linguistics*. Cambridge: Cambridge University Press.

Corbett, G.G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Corbett, G.G. 2012. *Features*. Cambridge: Cambridge University Press.

Harley, H. and Ritter, E. 2002a. "A feature-geometric analysis of person and number." *Language* 78 (3): 482-526.

Harley, H. and Ritter, E. 2002b. "Structuring the Bundle: A Universal Morphosyntactic Feature Geometry." In H. Simon and H. Weise (eds.) *Pronouns: Features and Representations*. Amsterdam: John Benjamin.

Macaulay, B.K. to appear. "The Morphosyntax of Formality: A Typology and Inclusion in Feature Geometry." *Penn Working Papers in Linguistics* 21 (1).

Nevins, A. 2007. "The representation of third person and its consequences for person-case effects." *Natural Language and Linguistic Theory* 25: 273-313.

Restrictions on co-occurrence of '*STEM + jānā*' and negation in Hindi: a contrastive analysis with '*-te + shimau*' in Japanese

Miki Nishioka
Osaka University
miki.nishioka@gmail.com

Shiro Akasegawa
Lago Institute of Language
lagoinst@gmail.com

1 Introduction

In this paper, we investigate restrictions on the co-occurrence in Hindi of the STEM form of the main verb plus the vector or auxiliary verb *jānā* 'GO' together with negative markers, using a Hindi corpus. It is well known that Hindi, a member of the Indo-Aryan language family, has a device of V1 (a main verb) + V2 (an auxiliary verb) to add a nuance of lexical aspect or modality to the meaning of the main verb. Masica (1991) labels meanings such as expressed by the V2s as *Aktionsart*. We can observe a quite similar device in Japanese, an East Asian language. Of the V2s used in Japanese, the verb *shimau* 'PUT AWAY' is frequently used to nuance the V1 in the *-te* form, i.e., one of conjunctive participles, in quite a similar way to *jānā* in Hindi.

This is one of the pragmatic issues that native speakers are only vaguely aware of, and thus find the usage of difficult to explain to non-native speakers – as the natives have already acquired competence of the target language and command it unconsciously.

2 Jagannathan's claim about the occurrence of V2s with negative markers

According to Jagannathan (1981), *jānā* 'GO', *lenā* 'TAKE', or *denā* 'GIVE', used as a vector verb, that is, as V2, never co-occur with negative markers like *nahiiN* or *na* in Hindi. Snell (2010), apparently in support of this claim, explains, "compound verbs give a specific sense of the way in which a particular action is done. It therefore follows that a sentence that's negative or general won't use them; …" [ibid, 290]. An illustrative example in Hindi is shown as follows.

[Hindi example]
(1) *kyā* *vo* *kitāb*
Q that book.F.SG
mil *gayī ?*
be found.STEM GO.PAST.F.SG
'Did you find that book?'

<Affirmative answer>

(2) *hāN, (vo kitāb)*
yes (that book.F.SG)
mil gayī.
be found.STEM GO.PAST.F.SG
'Yes, I did.'

<Negative answer>

(3) **nahīN, (vo kitāb) nahīN*
no (that book.F.SG) NEG
mil gayī.
be found.STEM GO.PAST.F.SG
'No, I did not find it.'

(4) *nahīN, (vo kitāb) nahīN*
no (that book.F.SG) NEG
milī.
be found.PAST.F.SG
'No, I did not find it.'

(2) is an affirmative answer to question (1). Theoretically, (3) is a negative answer, which consists of V1 + V2. Pragmatically, however, the proper negative answer is (4), which has lost the V2 of the original V1 + V2 sequences. Native speakers tend not to use answer (3), which although grammatically correct, is pragmatically unacceptable.

Looking at an example of *shimau* in Japanese, we can observe a quite similar phenomenon to *jānā* in Hindi, as follows.

[Japanese example]

(5) *shukudai (mō) yatte*
homework (already) do.CONJ-PART
shimatta?
PUT AWAY.PAST
'Have you finished your homework?'
[lit. 'Did you do the homework?']

<Affirmative answer>

(6) *un yatte*
yes do.CONJ-PART
shimatta.
PUT AWAY.PAST
'Yes, I have.' [lit. 'I did.']

<Negative answer>

(7) **iya (mada) yatte*
no (yet) do.CONJ-PART
shimawa-nakatta.
PUT AWAY.STEM-NEG.PAST
'No, I have not.'
[lit. 'I did not put (doing) it away (yet).']

(8) *iya (mada) yatte*
no (yet) do.CONJ-PART
inai.
COP.STEM-NEG.NON-PAST
'No, I have not.'
[lit. 'I have not done it (yet).']

Question (5) with *shimau* as a V2 is answered in the affirmative as in (6). However, for negative answers, the sentence with *shimau* and the negative marker as in (7) is hardly acceptable. The appropriate answer should be (8), where *jānā* as a V2 is replaced with the Japanese copula *iru*. This kind of phenomenon may prove that although Hindi and Japanese are said to be from different language families, both have the same devices, which belong to the same or to very similar grammatical categories, even though they utilize different lexical verbs.

3 Do V2s really not occur at all with negative markers in Hindi?

So does a V2 never occur with a negative marker? Jagannathan (ibid) has claimed that there are some exceptions in Hindi. These are as follows:

[Hindi]

(9) *tum so na*
you sleep.STEM NEG
jānā.
GO.INF (IMP)
'Do not sleep.'

(10) *jab tak sāre log andar*
when till whole people inside
na ā jāẽ...
NEG come.STEM GO.SUB
'not until all of people come inside, ...'
[lit. until when not all of people come inside,]

The infinitive form as a finite predicate in (9) is used in the imperative. (10) is also an exception. Both are appropriate sentences. Jagannathan himself pointed out that the V2s in this kind of sentence are 'necessary' or 'mandatory', not 'optional'. However, this rule is somewhat unclear, especially to non-native Hindi speakers. To illustrate the difficulty, we shall analyze Jagannathan's point in detail, focusing on examples gathered from large-scale corpora.

4 What do corpora tell us about co-occurrence restrictions?

Based on the above and on the hypothesis that *jānā* in Hindi and *shimau* in Japanese as a V2 belong to very similar grammatical categories and behave in a similar way, we have collected examples of *-te* form + *shimau* sequences, compiled statistics on the co-occurrence of the auxiliary *shimau* and negative markers, and examined in what circumstances these examples are used - that is, what kinds of elements precede and follow negated *-te* form + *shimau*

sequences, using a corpus called The Balanced Corpus of Contemporary Written Japanese (BCCWJ). What we have found is that: the number of negative *-te* form + *shimau* sequences we've received from the corpus was quite small compared to the number of the affirmative *-te* form + *shimau* sequences; the results of the negative examples were not found in indicative sentences; all of the examples in negatives are restricted to expressions that are equivalents of moods, such as imperatives, subjunctives and conditionals. or to modalities that express a speaker's mental state or judgment vis-a-vis the propositional content of an utterance – which leads to Jagannathan's point on 'mandatory' vs. 'optional'.

We deduce that most of the facts we have found in the Japanese corpus will be applicable to *jānā* as a V2 in Hindi. With respect to Hindi corpora and examples, we have built an experimental corpus of collected data from websites, which has about 10,000,000 words. The main objective findings observed in our results from the Hindi corpus are as follows:

- The frequency of occurrence of negative *-te* form + *shimau* sequences is smaller than that of of affirmative ones.
- When used in the indicative mood, the Hindi *jānā* hardly ever occurs with a negative marker.
- The *jānā* occurs with a negative marker when used in the imperative.
- There seem to be no restrictions on using *jānā* as a V2 with a negative marker in adverbial clauses such as conditional, subjunctive, or adjective clauses or noun clauses.

All of the findings observed above are identical to our findings from the Japanese corpus.

5 Conclusion

As far as we have observed from the corpora, we can ascertain, firstly, that V2s like *jānā* and *shimau* are in nearly the same grammatical category, although they are from non-cognate languages. Secondly, based on this fact, we can also point out that the V1 + V2 concatenation behaves as a kind of lexicon. In other words, *jānā* as a V2 itself cannot be negated, whereas examples other than the indicative show that the negative markers belong not to the proposition (the V1 + V2 concatenation), but rather to some extraneous semantic element such as an imperative, subjunctive, conditional, etc.

References

Hook, P.E. 1974. The Compound Verb in Hindi. Michigan: University of Michigan, Center for South and Southeast Asian Studies.

Jagannaathan, V.R. 1981. *Prayog aur prayog.* Delhi: Oxford University Press.

Kachru, Yamuna. 1980. Aspects of Hindi Grammar. New Delhi: Manohar.

Kellogg, H. Samuel. 1938. A Grammar of the Hindi Language : in which are treated the High Hindí, Braj, and the Eastern Hindí of the Rámáyan of Tulsí Dás, also the colloquial dialects of Rájputáná, Kumáon, Avadh, Ríwá, Bhojpúr, Magadha, Maithila, etc., with copious philological not (the 3rd edition). London: Kegan Paul, Trench, Trubner and Co.

Masica, C.P. 1991. *The Indo-Aryan Languages.* Cambridge: Cambridge University Press.

National Institute for Japanese Language and Linguistics and Lago Institute of Language, 2012. *NINJAL-LWP for BCCWJ (NLB)*, NINJAL, Tokyo, accessed 5 October 2012, < http://nlb.ninjal.ac.jp/>.

Nishioka, Miki. 2013. "Te-kei + *shimau* to hiteiji to no kyōki seigen to sono kankyō ni tsuite: Hindīgo to no taishōgengogakuteki shiten kara (Co-occurrence Restrictions on the '-*te* Form + *shimau*' and Negation in Japanese: A Contrastive Analysis with Hindi)", *Matani ronshū vol.7*. Pp.47-73, Osaka: Nihongo Nihon Bunka Kyōiku Kenkyūkai.

McGregor, R.S. 1995. Outline of Hindi Grammar. Oxford: Oxford University Press.

Snell, Rupert. 2010. Teach Yourself Get Started in Hindi (Teach Yourself Beginner's Languages). 2nd Revised. Teach Yourself Books.

Copy-verb constructions in colloquial Burmese

Pavel Ozerov
La Trobe University
p.ozerov@latrobe.edu.au

1 Introduction

This study analyses a peculiar verbal pattern, dubbed a "copy-verb construction", in colloquial Burmese on the basis of a corpus of written and spoken language. This construction consists of a sequence of two verbal forms: a non-finite verbal stem and the repetition of this verb in its finite form. The non-finite constituent is marked for its discourse structuring or information structural role by a designated particle. The finite verb serves as the locus of the obligatory verbal categories, e.g. TAM, as in (1).

(1) *di-kãu-ká* ***pjò=θa***
this-guy-SBJ **speak=only**
pjò-já-ta
speak-NVL-R.NMLZ
'This guy is just talking! (i.e. he is bragging, but cannot do anything)' [Thadon]

Although typologically these constructions are not uncommon (they are known from European, Semitic, African and other languages), allusions to them in the literature are mainly sporadic and often their analysis proves to be troublesome. Hyslop (2011: 681) briefly describes a parallel syntactic construction with "emphatic" particles in Kurtöp, while Konnerth (2014: 586–592) finds a set of similar constructions in Karbi that have an intensifying role. However, it appears that copy-verb constructions have not been extensively addressed in the literature (see Daudey 2014 for the only study dedicated to this issue).

2 Copy-verb construction in Burmese

A verb in Burmese represents a single morphosyntactic whole that includes the verbal stem and a variety of markers of TAM and other categories. From the point of view of information structure, it is usually assumed that this complex corresponds to a uniform status, typically the focus. Assigning a topical status to a verb requires special strategies like nominalisation. Yet, in both cases all the categories "trapped" within the verbal conglomerate (TAM, polarity) have joint information structural and discourse contribution with the lexical constituent (stem). However, this is not always the case and some contexts call for a decomposition of the verb and a separate assignment of informational roles to its internal constituents.

Analysis of corpus data from spoken and written colloquial Burmese shows that the construction is employed for this kind of information decomposition of the verbal complex with the purpose of a separate pragmatic manipulation of its constituents. Since the only relatively independent component of the verb is its lexical stem, the desired interpretational outcome is achieved through extra-position of the stem and its marking by "pragmatic" particles. Other verbal categories like TAM and polarity are expressed through clitics/affixes to the stem, and hence require an overt lexical stem to attach to. This results in its repetition and in the copy verb construction as the final outcome.

The copied part of the construction is the first syllable of the verb. If a verb is not monosyllabic, it is the first syllable that is copied. In the case of a bi-syllabic lexical compound that includes a noun as its first constituent, it is this noun that is copied. This is the case in (2): literally, *wǎ* means 'stomach', *θa* – 'pleasant' and *nɛ̀* – 'little'. 'To be happy' and 'to be sad' are lexicalised expressions *wǎ-θa* ('stomach-pleasant') and *wã-nɛ̀* ('stomach-small') accordingly. It can be seen that only *wǎ* 'stomach' is pre-posed in (2), although it does not contribute its lexical meaning directly, but forms a part of the lexicalised compound.

(2) ***wǎ=lɛ̀*** ***wǎ.θa-tɛ,***
COPY=ADD **happy**-R,
wǎ=lɛ̀ ***wǎ.nɛ̀-tɛ***
COPY=ADD **sad**-R
'I am both happy and sad.' [ThA0614_AMT]

When the information conveyed by the verbal stem is strongly suggested contextually, it can be straightforwardly introduced as a new topical constituent. In this configuration it is expressed as a copied element with the marker *tó* that typically triggers an interpretation of a new or contrastive topic.

(3) An interlocutor interrupts a story about a grave wound, received by a living person, to make sure his understanding of the final outcome is correct (the issue of the wound closing up is obviously accessible and highly relevant).
pei?=tópei?-θwà-tɛ=pó
close=TOP **close**-GO-R.NML=RINF
'**To close up** – it finally **did** close up, (right?)' [ThA0614_Hun2]

Yet, topicalisation of the verbal stem additionally results in the transferral of the focal burden on the other categories of the verb, creating a focal interpretation on some of them, like the polarity.

Both effects arise in (3) simultaneously. Moreover, when the verbal information is contextually redundant, focalisation of a category like polarity is the only motivation for its usage, as in (4).

(4) 'Women can **have** feelings of embarrassment, shame, guilty conscience. They **have** this doubting mind. It may differ – some **have** it more, some **have** it less.'

ɕí=t́ó *ɕí-tɕá-t́ɛ*
have=TOP **have**-PL-R
'But **to have**$_{TOP}$ it – they *do* **have**.' [Blog 2]

Alternatively, the verbal lexeme can require a focal interpretive effect. In this case the copied verb is followed by one of the focus-markers available in Burmese. For instance, this is the case in (1), where the marker *θa* has the exclusive-derogatory interpretation reminiscent of the English 'just'. The marker *tãu* often receives an interpretation comparable to that of 'even': it conveys that the asserted element is to be contrasted with the previously suggested options and represents the less expected edge candidate, as in (5).

(5) The wolf accused the lamb in insulting him during the previous year. The lamb answers:

mə.n̥iʔ-tôu-ḵá *tɕənɔ* ***mwè=t́ãu***
last.year-period-ABL 1 **be.born=even**
*mə-**mwè**-θ̱è-p̱a-p̱*^{h}*ù*
NEG-**be.born**-yet-POL-NEG$_2$
'Last year I was not **even** born yet.' [Fables 13]

Unlike the contrast for other possibilities obtained in (5), the marker *n̥á* in (6) triggers the interpretation of the extreme degree within the state of affairs expressed.

(6) ***jõ=n̥á*** *mə-**jõu**-bɛ̀=nɛ̀*
believe=**FOC** NEG-**believe**-NF.NEG$_2$=with
'I did not believe ***at all*** and...' [ThA0614_Hun]

This kind of focal interpretation is unknown in the studies of information structure in European languages, where the corresponding translation typically requires a usage of intensifiers (e.g. 'at all'), whose contribution is primarily regarded as discourse-"emphatic", rather than information-structural. Interestingly, Burmese employs grammatical focus-marking for the expression of this category.

Finally, copy-verb construction is used with the marker *lɛ̀* for cumulative/additive discourse coordination of multiple propositions ('and, also' etc.). Typically *lɛ̀* attaches to the initial nominal constituent, as in (7)(b). However, if such an element is unavailable because the clause consists of a single verb, as in (7)(a), the first syllable of the verb is copied to host the clitic.

(7) The speaker outlines the advantages of the spectacle chosen for a show.

(a) ***tɕì=lɛ̀*** ***tɕì**-tɕɛ-já-mɛ*
be.big=ADD **be.big**-be.loud-NVL-IRR
'It will be great.'

(b) *təjà-θãwegá=**lɛ̀*** *já-já-mɛ*
moral-learning=**ADD** get-NVL-IRR
'And one can learn morals [from it].' [Thadun]

3 Conclusions

Burmese employs copy-verb construction to decompose the morphosyntactic verbal complex and assign information structural roles separately to its constituents. This is done through the manipulation of the status of the verbal stem. This stem can acquire topical interpretations as a result of the marking *t́ó*, with ensuing focal interpretations of the remaining categories like polarity. Alternatively, it can receive focal interpretation of exclusiveness, excessiveness etc. due to an according marking.

The construction bears remarkable cross-linguistic parallels with other reminiscent structures and corresponding functions in various Tibeto-Burman languages like Wadu Pumi (Daudey 2014), Karbi (2014: 586–592), Kurtöp (Hyslop 2011: 681). A coordinating function can be found e.g. in Numphuk Singpho (Morey 2011: 295).

References

Daudey, Henriette. 2014. Copy-verb constructions in Wadu Pumi. Paper presented at ICSTLL 47, Yunnan Normal University, 18 October

Hyslop, Gwendolyn. 2011. 'A Grammar of Kurtöp'. Eugene: University of Oregon, PhD dissertation.

Konnerth, Linda. 2014. 'A Grammar of Karbi'. Eugene: University of Oregon, PhD dissertation.

Morey, Stephen. 2011. in Yap, Foong Ha, Karen Grunow-Hårsta & Janick Wrona (eds.) *Nominalization in Asian Languages: Diachronic and Typological Perspectives*. 289–312. Amsterdam/Philadelphia: John Benjamins.

The origins of the New Indo-Aryan postpositional phrases

Uta Reinöhl
University of Cologne
uta.reinoehl@uni-koeln.de

The origins of the New Indo-Aryan postpositional phrases, as exemplified by Hindi *choṭe kamre mẽ* 'in a/the small room' or *Rām ko* 'to Rām' have been traced to an appositional construction consisting of the respective etymological forerunner (e.g. Skt. *upari* 'above' being at the origin of Hindi *par* 'on') and a local case form in strings such as *upari parvate* 'above, at the mountain' (cp. Bubenik 2006; Chatterji 1970 [1926]; Coleman 1991:232; Emeneau 1956: 9). It is claimed that, as a result of frequent co-occurrence in juxtaposition, such combinations were at some point reanalyzed as syntactic units, i.e. as the postpositional phrases that we find today. I present evidence for an alternative origin that traces the postpositional phrase not to appositional constructions, but to asymmetric structures consisting of a nominal or verbal head together with an adnominal genitive or dependent direct object (in different case forms) respectively.

Adpositional phrases (specifically such phrases containing "simple" adpositions as e.g. English *of*, *on* or French *à*, *de*) of modern Indo-European languages are in general assumed to have arisen from appositional constructions involving spatial adverbs. In most branches, these are the so-called local particles, a class of elements that survive not only as adpositions, but also as preverbs in many daughter languages. For example, Latin *ad* 'at, to' is reconstructed to a Proto-Indo-European local particle *$*h_2ed$* and survives in French both as an adposition *à* 'at' and as a preverb, e.g. in *arriver* 'to arrive'. Co-occurring with accusatives, for instance, the string was reanalyzed yielding adpositional phrases such as *ad Romam* 'to Rome'. Cognates of these local particles were also present in Vedic Sanskrit. In Vedic, they can still be regarded a single class of adverbial elements syntactically, but semantically often already modify either verbal or nominal forms. The New Indo-Aryan postpositions, however, do not derive from these local particles etymologically. This well-known fact has not been considered a fundamental difference between Indo-Aryan and other branches. Instead, it is assumed that the NIA postpositions go back, if not to the same local particles (these vanish as free morphemes at an early stage in Sanskrit and only survive as preverbs in Indo-Aryan), then to other similar types of spatial adverbs. Data from Vedic Sanskrit indeed suggests that this scenario is correct, as the etymological forerunners of the modern postpositions frequently or exclusively occur as adverbs, often in juxtaposition with another local case form. This goes both for postpositions deriving from nominal forms (e.g. Hindi/Urdu *mẽ* deriving from *madhye*, middle.LOC, 'in the middle') or from adverbs (e.g. *par* deriving from *upari*). (I disregard here elements such as Hindi/Urdu *kā/ke/kī*, for which a participial origin can be assumed.) Thus far, we seem to be dealing only with a difference in etymons, but not with a different origin of the adpositional phrases on a categorical level (abstracting from the difference in word order).

If, however, one takes into consideration evidence from younger stages of Sanskrit, as well as from early Middle Indo-Aryan, the picture changes. Several elements that used to occur only or almost exclusively in adverbial function now begin to appear with adnominal genitives (cp. Bloch 1965: 159). For example, while *upari* occurs in Vedic exclusively as an adverb, it can take genitives from Classical Sanskrit onwards. In Pali, this usage is more frequent than its old adverbial usage. This new construction is characterized also by a different semantic structure. In apposition with local case forms, *upari* means 'above', i.e. denoting the upper area in a three-dimensional space. With genitives, by contrast, it means 'on', e.g. Pali *payasotattassa* (GEN) *nibbāyamānassa* (GEN) *upari* 'on boiled milk' (Dīgh III, p. 85). In later stages of Middle Indic, constructions with genitives begin to outnumber constructions with local case forms by far. This goes both for constructions with an originally nominal head as e.g. *madhye*, which could always take adnominal genitives, as well as with a head that used to be adverbial, e.g. *upari*, and only acquired the capacity of taking adnominal genitives. Eventually, only constructions with genitives remain, the latter merging into the general oblique forms that we find today. The evidence suggests that it is these asymmetric constructions that are at the origin of the modern postpositional phrases. There are additional indications from semantics. Some usages of the modern postpositions can be traced only to constructions with genitives. For example, Hindi/Upari *par* can be used to express a feeling towards someone, similar to German *auf* ('on') *jmd. böse sein* 'be angry at someone'. This usage has predecessors that can be found as early as in Pali in constructions with genitives, but is unattested in combinations with local case forms.

How to account for this reanalysis of adverbs as syntactically nominal, allowing for adnominal genitives? First of all, there was a strong analogical model, namely relational noun constructions of the type GEN + *madhye*, GEN + *samīpe*, GEN + *antare* etc. Second, only such adverbs (e.g. *antarā*

'between' or *saddhiṃ* 'with' besides *upari*) undergo the reanalysis, which have a phonological shape that can be taken as a local case form, e.g. *upari* as a locative of an *r*-stem. Third, due to expanding case syncretism in younger stages of Sanskrit and especially Middle Indo-Aryan, we often cannot know whether we are dealing with a local case form or with a genitive encoding the spatial reference frame.

In sum, based on a corpus study of Old and Middle Indic data carried out in Reinöhl (forthc.), I propose that the postpositional phrases of New Indo-Aryan must be traced to asymmetric constructions involving nominal(ized) (and verbal) heads and their respective dependents, rather than from appositional constructions. This conclusion can only be drawn if several historical stages are taken into consideration. In particular, an exclusive look at Vedic will lead to incorrect conclusions. In general, one cannot assume that Indo-Aryan parallels the developments in other branches of Indo-European, even if the situation in Vedic appears to be largely parallel to that in other ancient varieties. It becomes clear that the historical developments of Indo-European are not as homogeneous as has been assumed both in classical (e.g. Meillet & Vendryes 1927) and in more recent studies (e.g. Luraghi 2010). Above and beyond questions particular to the language family, the evidence from Indo-Aryan shows that there is more than one way in which nominal expressions in a strongly non-configurational language may come to be phrasally organized.

References

Bloch, J. 1965. *Indo-Aryan. From the Vedas to modern times*. (revised and translated by Alfred Master). Paris: Librairie d'Amérique et d'Orient Adrien – Maissonneuve.

Bubenik, V. 2006. "Cases and Postpositions in Indo-Aryan", In J. Hewson & V. Bubenik (eds). 2006. *From Case to Adposition. The Development of configurational syntax in Indo-European languages*. Amsterdam/Philadelphia: Benjamins, 102-130.

Chatterji, S. K. 1970 [1926]. *The Origin and Development of the Bengali Language*. London: Allen & Unwin.

Coleman, R. 1991. "Latin prepositional syntax in Indo-European perspective", In: R. Coleman (ed.). *New Studies in Latin Linguistics*. Amsterdam/Philadelphia: John Benjamins, 323-338.

Emeneau, M. B. 1956. "India as a Linguistic Area", *Language* 32 (1), 3-16.

Luraghi, S. 2010. "The rise (and possible downfall) of configurationality", In S. Luraghi & V. Bubenik (eds). *The Continuum Companion to Historical Linguistics*. London/New York: Continuum, 212-229.

Meillet, A. & J. Vendryes. 1927. *Traité de grammaire compare des langues classiques*. Paris: Champion.

Reinöhl, U. Forthc. *Grammaticization and Configurationality – The Emergence of Postpositional Phrases in Indo-Aryan.* Oxford: Oxford UP.

Are Post-Nasal Plosives the Proto Form of Dravidian Language?

Honnappa Shashikala
Bangalore University
shashih5@yahoo.co.in

In the Reconstruction of Proto language, Comparative method of the languages of one family is followed. While comparing the languages for the purpose of reconstruction, we should consider all cultivated and un-cultivated languages. Because, many variations will be available and also these variations are important in the reconstruction of Proto-Language of one family.

Variation in cognates may be telling different stories with regard to the history of particular language. Variation in cognates may be due to the changes in language. Variation in cognates must be observed very keenly, because there might be secondary developments, and may not belong to the same stage.

Sound changes are regular or irregular. Descriptive methodology is used hitherto to describe sound changes. But sound changes are to be explained with new methods. Because we have to search for the reasons for these sound changes. So it is important to take note of changes in the utterance form at different stages.

In Dravidian Languages, formative suffixes play a major role in the sound change. There is direct relationship between sound change and changes in the syllabic structure. Changes in the syllabic structure will lead to sound change also. Sound changes should be explained through Morphophonemics .

A Proto-Dravidian (PDr.) root is monosyllabic and is either open or closed. "In many cases, the heavy base (C) V̄C- can itself be etymologically resolved into $(C)^1V^1$-C^2-, where –C^2- is part of a derivative; but this development should be attributed to the period preceding the common PDr Stage" (Zvelebil .K, 1970)

Eg.,*kal-u. Here –u is enunciative vowel. But such type of enunciative vowels also play the role of formative suffixes. These formative suffixes will play a major role in sound change in any language.

Here we consider the alternation occurred in the postvocalic geminate retroflex consonant (PP) and the nasal-retroflex consonant cluster (NPP) in Kannada, by taking the example of tōṭa (Garden). The reconstructed Proto-form for tōṭa (garden) is tōnṭṭam (NPP). whether this proto-form which has been reconstructed by the linguists holds good or not is the research of this article.

The data which is available in the inscriptions of Kannada Language is given here;

I /tōṭṭ-a/ (5th century inscriptions)
II /tōṇṭ-a/ (8^{th} century inscriptions)
III /tōṭ-a/ (later inscriptions)

Other Dravidian Languages, particularly uncultivated languages show both PP and NP structures. There are ample evidences showing gemination changes (PP) to Nasals plus Plosives (NP). This must be intermediatory or secondary stage in the history of Dravidian Languages. Before Deleting the Nasals or Changing Voiceless to Voiced, this intermediatory stage - NP/NPP existed.

1. Eg. ṅk. (<PP=kk) (DED-37)
   Ta. ankati 'bazaar',
   Ma. annati,
   Ko. angady,
   To. Ogady
   Ka. aṅgaḍi

   Correspondences are:
   Ta. aṅkā 'to open' (DED-36)
   Ta. **aka**l 'to spread'(DED-9)

2. Eg. Ka. aṇṭu, aṭṭu- to stick, adhere to; (DED 86)
   Tu. aṇṭu, aṇṭu pattu, aṇṭāvuni
   Te. aṇṭu; aḍug- aṇṭu
   Ta. aṭṭu
   Nk. **aṭṭ**
   Kol. **aṭ-**

Above data shows the postvocalic geminate retroflex consonant (PP) and the nasal-retroflex consonant cluster (NPP) in the different languages of Dravidian Family. By considering these data, scholars have reconstructed the postvocalic geminate retroflex consonant (*PP=ṭṭ) and the nasal-retroflex consonant cluster (*NPP=ṇṭṭ) and retroflex consonant (*P=ṭ), were the Proto-Dravidian forms. And also from nasal consonant gemination changes to consonant gemination.(NP>PP/P)

These reconstruction were done by Kumaraswami Raja (1969) and accepted by various scholars, among them are Bh. Krishnamurti(2003), DNS Bhat(2007), P.S.Subramanayam(1983).

But the changes occur in different way. Gemination changes to nasal, and then in the next stage, nasal deletes. Before deletion of gemination, one plosive weakens and changes to nasal. Actually here, in the understructure, plosive turns into nasal.

Some of the Dravidian languages underwent this stage; some of them did not. The scholars while reconstructing proto form consider nasals + plosives and germination of plosives both as proto form . Because of this variation Kumara swami Raja could not specifically pinpoint the pattern of the change.

Kumara swami Raja (1969) states: "Instances where the Telugu-group of languages show PP as a reflexes of *NPP are less numerous than those which show the regular reflex. No explanation is possible for this discrepancy. Only future research can throw light on this."

In my paper, I am proposing that geminate plosives are the proto form of Dravidian languages, and not post-nasal plosives. So Reconstruction of proto-form here in such cases is possible only with geminate plosives.

References

N.Kumaraswami Raja(1969), Post-Nasal Voiceless Plosives in Dravidian, , Annamalai University,Annamalainagar

Bh.Krishnamurti (2003),*Dravidian languges*, Cambridge Universty Press

T.Burrow and M.B.Emeneau (1984), *Dravidian Etimological Dictionary*, Munshiram Manoharlal, New Delhi.

Andrew Radford, Martin Atkinson, David Britalin, Harald Clahsen, Andrew Spencer,(2009), *Linguistics an Introduction*, Cambridge Universty press, UK

K.V.Narayana,(2011), *Kannaḍada ulikanthegaḷu*, Pragati Graphics, Bengaluru

DNS Bhat (2007), *Kannaḍa nuḍi naḍedu banda haadi*, Bhasha Prakashana, Sagara.

P.S.Subramhmanyam, (1983), Dravidian Comparative Phonology, Annamalai University,Annamalainagar

M.B.Neginahala (1982), Prācheena Kannaḍa Śasanagaḷa Bhaśika adhyayana, Karnataka University, Dharwad

The diachrony of cosubordination – lessons from Indo-Aryan

Krzysztof Stroński
Adam Mickiewicz University in Poznań
stroniu@amu.edu.pl

Joanna Tokaj
Adam Mickiewicz University in Poznań
joannatokaj@wp.pl

1 Introduction

The notion of cosubordination plays a central role the theory of the Role and Reference Grammar (RRG) in defining nexus types (e.g. Van Valin and La Polla 1997). Cosubordination is an abstract linking relation which assumes operator dependency and non-embededness (cf. Van der Auwera 1997).

The question if the operator sharing is a defining feature of the third linking type was addressed in the typological literature and Bickel (2010) proposed a more fine-grained solution advocating a number of structural parameters which should bring us closer to the problem of linking types other than subordination and coordination. Here we are going to focus on the converbal chain construction utilizing partly a set of variables which was proposed within the framework of the multivariate analysis (Bickel 2007; 2010).

We first briefly define the notion of cosubordination, then we give the main construction types which are supposed to be instances of cosubordination in IA according to the metrics proposed by RRG and evaluated later by the multivariate analysis, next we discuss the main constructions diachronically showing the behaviour of main parameters and finally we draw some preliminary conclusions.

2 Cosubordination

RRG assumes that there are three basic nexus types i.e. coordination, subordination and cosubordination (cf. Van Valin & La Polla 1997: 454; Van Valin 2005: 187-188). Cosubordination differs from coordination and subordination in the following way. On one hand, it does not involve embedding e.g. cosubordinative clauses are neither modifiers nor arguments of the clauses they are linked with and this brings the relation closer to coordination. On the other hand the cosubordination is a dependency relation but in contradistinction to subordination where the dependency is structural here it is operator dependency.

Cosubordination operates at all three levels of juncture, namely nuclear, core and clausal. At each juncture level there are operators from which at least one should be shared (Van Vallin 2005: 201). Thus

we have nuclear operators (nuclear directional, nuclear negation, aspect), core operators (modality, core directionals, internal negation) and clausal operators (tense and illocutionary force). Operator sharing is hierarchized i.e. operator sharing at a certain level of juncture assumes sharing at higher level operators as well but the reverse is only possible if it does not violate the semantics of the predicate (Van Valin and La Polla 1997: 455).

3 Converbal chains vs. cosubordiantion

The construction labelled converbal chain has in its centre a non-finite verbal form, i.e. converb which has been widely discussed in the literature from typological, descriptive or historical perspectives (for extensive survey see e.g. Haspelmath and König 1995). In the Indo-Aryan scholarship the notion of converb occupies a special place since it is one of the elements defining a so called 'linguistic area' (cf. Masica 1976) and the form itself has been analysed at various stages of IA and in a number of dialects (e.g. Dwarikesh 1971; Schumacher 1977; Davison 1981; Kachru 1981; Tikkanen 1987; Peterson 2002; Yadav 2004; Lohar 2012). There have been also typological and areal studies devoted to the problem of converb in IA (cf. Masica 1976; Subbarao 2012).

From the RRG perspective, converbs can be a part of constructions which instantiate different types of cosubordination, namely nuclear, core and clausal. We focus here exclusively on the clausal junction. For example in (3) the first clause consist of a converbal core and the second one has a finite verb. The IF operator has scope over the two clauses. Since both clauses are interpreted as commands this is an instance of a clausal cosubordination.

The border between core junction and clausal juncture is not clear-cut but there is a semantic difference between the converbal core of the sort illustrated in (1) and in (2), namely the clause-like character of the former and the adverbial character of the latter. Therefore only the former is a converbal chain construction.

1. Hindi
to *tumhī* *jā-kar* *dekho* *na?*
CONJ you.OBL.EMPH go-CVB see.IMP.2PL not
'So you go and see, OK?' (Premcand, *Kafan*, 8)

2. Punjabi
kuṛī *muskā-ke* *bol-ī*
girl.F.SG smile-CVB spoke.PAST.F.SG
'The girl spoke smilingly.'

Clausal junctions involving converbs seem to be one of the most important linking devices not only in IA but generally in South-Asian languages. At first look the converbal chains instantiate cosubordiantion as we see in (1) but it has been demonstrated in the literature that the IF operator does not have to be necessarily shared by the two clauses as it always is in the case of clausal cosubordination (cf. Peterson 2002; Bickel 2010).

3. Nepali
phalphul *tāch-era* *nānī-lāī* *di-ūṂ?*
fruit peel-SEQ little.girl-OBJ give-INJ.1S
'After I peel the fruit, should I give it to the child?'
'Should I peel the fruit, before I give it to the child?'
'Should I peel the fruit and give it to the child?'
(Peterson 2002: 105)

4 Methodology

Structures of the type (3) led Bickel (2010) to a conclusion that the notion of cosubordination is not easily definable via the IF operator sharing and he proposed a scalar approach by means of which it was made possible to place structure of this particular type between subordinative and coordinative ones. The list of variables proposed in the 'multivariate analysis' (Bickel 2010) is a lengthy one and here we focus only on those which are most relevant to the constructions under consideration. The variables presented here have been adopted from Bickel (2006; 2010):

a) Scope of the IF operator : conjunct, disjunct, local, extensible, constraint-free
b) Tense Scope: conjunct, local, extensible
c) Who-questions in dependent clauses: banned, allowed
d) Focus marking in dependent clauses: banned, allowed.

Firstly we are going to demonstrate how the system works for the converbal chain construction in the text corpus consisting of early Rajasthani prose texts from the 15th to 17th centuries counting approximately 10000 words. Then the early Rajasthani data will be compared with other early NIA dialects such as Braj, Awadhi for which the corpora of approximately the same size have been inspected[14].

5 Converbal chains in early NIA: preliminary results

Preliminary research shows that the scope of illocutionary operators can be local (limited to the main clause) or conjunct. This is further confirmed

[14] Rajasthani texts have been excerpted from Bhānāvat and Kamal (1997-1998), Awadhi from Gautam (1954) and Braj from Vājpeyī (2009) and they have been tagged by means of the 'IA tagger' – a programme designed for tagging early NIA texts at 5 different levels, namely – morphological, parts of speech, syntactic, semantic, pragmatic (IA Tagger 2014). Optical recognition of Rajasthani texts was supported by a Hindi OCR programme (HindiOCR 2013).

in contemporary NIA, cf. Davison (1981), Bickel (2010)). The examples below show the local or conjunct scope of imperatives in early Rajasthani (with a clear preference for the conjunct scope) from the 16th and 17th centuries (4) and (5) and conjunct scope in early Awadhi (6):

4. early Rajasthani, 16th century
paṇi tumhẽ mayā karī deśāntari pahucaü
but you mercy do.CNV abroad.OBL reach.IMP
'But you, having shown mercy, go abroad'.
'But you show mercy and go abroad'. (R.G. 33)

5. early Rajasthani, 17th century
upāṛi-ara ghoṛā māṃh ghālo
lift-CVB horse.NOM.M.PL in throw.IMP
'After lifting the horses throw them into (the river).'
'Lift the horses and throw them into (the river).' (R.G. 44)

6. early Awadhi
kai cali hohu suā sanga satī
or go.CNV be.IMP parrot.M.SG with satī.F.SG
... or go and become satī with the parrot. (J.88.8)

Tense operators in early NIA had predominantly conjunct scope (7) although this is not always the case in the contemporary IA (cf. Peterson 2002).

7. early Rajasthani, 17th century
jyūṃ utari pāri jāhiṃ
as get out.CVB cross.CVB go.SBJ.3PL
'If (they) get out they will cross (the river)'.
(R.G. 44)

Question operators in early NIA had basically local scope (see (8) from early Braj), although in contemporary NIA they can have conjunct scope as well (cf. Davison 1981; Bickel 2010). The same pertains to negation (9). It seems that the scope should be dependent on the position of the question or negation markers, but this still requires more detailed cross-linguistic verification.

8. early Braj
aü so gāruri māṅgi kaï
and he snake charmer request.CVB what
līnhā
take.PPP.masc.sg.
And having requested snake charmer what did he take? ('Lorakahā' 53 15th c. from Tripāṭhī 1957: 124)

9. early Braj
bhate haṅkāri na pūhāhu tāhī
hero.OBL call.CNV not ask.IMP.2PL he.OBL
Having called a hero do not ask him. ('Haricarita' 1.12 15/16th c. from Tripāṭhī 1957: 127)

The corpora inspected by us show also quite remarkable differences between converbs and past adverbial participles not only at the syntactic but at the pragmatic level as well. Question operators seem to have basically local scope whereas tense operators have never conjunct scope.

Our preliminary research attempts at outlining general properties of selected non-embeded structures in early NIA showing that presumably so called cosubordination is not a discrete notion but rather a bundle of features which have been quite stable over centuries.

References

Bhānāvat, N. and Kamal, L. (eds.). 1997–1998. *Rājasthānī gadya: vikās aur prakāś*. Āgrā: Śrīrām Mehrā end Kampanī. (R.G.)

Bickel B. 1999. "Grammatical relations, agreement, and genetic stability". Ms., University of California, Berkeley.

Bickel, B (2010). "Capturing particulars and universals in clause linkage: a multivariate analysis". In: Bril, I. (ed.) *Clause Linking and Clause Hierarchy : Syntax and Pragmatics*. Amsterdam, 51 - 102.

Davison, A. 1981. "Syntactic and semantic indeterminacy resolved: a mostly pragmatic analysis for the Hindi conjunctive participle". In: Cole Peter (ed), 1981. Radical pragmatics. New York: Academic Press, 101–128.

Dwarikesh, D. P. 1971 . *Historical syntax of the conjunctivc participle phrase in New Indo-Aryan dialects of Madhyadesa (Midland) of northern India*. (University of Chicago Ph.D . dissertation).

Gautam, M. 1954. *Jāyasī granthavalī*. Dehli: Rigal Buk Ḍipo. (J.)

Haspelmath, M. and König, E. (eds.), 1995. *Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms – adverbial participles, gerunds*. (Empirical approaches to language typology, 13.) Berlin: Mouton de Gruyter.

Kachru, Y. 1981. 'On the Syntax, Semantics and Pragmatics of the Conjunctive Participle in Hindi – Urdu'. *Studies in Linguistic Sciences* 1981, Volume 11. Number 2, 35-49.

Lohar, G. Th. 2012. "Converbal constructions in Bhojpuri". *Nepalese Linguistics* 2012, Volume 27. 217-222.

Masica, C. P. 1976. *Defining a linguistic area. South Asia*. Chicago and London: Chicago University Press.

Peterson, J. 2002. "The Nepali converbs: a holistic approach." In: Rajendra Singh (ed.): *Yearbook of South Asian Languages and Linguistics*. New Delhi/Thousand Oaks/London: Sage Publications. 93-133.

Schumacher, R. 1977. *Untersuchungen zum Absolutiv in modernen Hindi*. Frankfurt a.M.: Lang.

Subbārāo, K. V. 2012. *South Asian languages: A Syntactic Typology*. Cambridge: Cambridge University Press.

Tikkanen, B. 1987. *The Sanskrit gerund: a synchronic, diachronic and typological analysis*. Studia Orientalia, 62, Helsinki.

Tripāthi, V. 1972. *Prārambhik avadhī kā adhyayan*. Ilāhābād: Racnā Prakāśan.

Vājpeyī, S. (ed.), 2009. *Sūrsāgar*. Kāśī: kāśī nāgarī pracāriṇī sabhā.

Van Valin, R. & LaPolla, R. 1997. *Syntax*. Cambridge: Cambridge University Press.

Van Valin, R. 2005. *Exploring the Syntax–Semantics Interface*. Cambridge: Cambridge University Press.

Yadav, R. 2004. "On diachronic origins of converbs in Maithili". *Contributiions to Nepalese Studies*, Vol. 31, Num. 2, 215-241.

Encoding motion events in Kannada: Is it in English way or Spanish way?

N. P. Sudharshana
Dept. of Humanities and Social Sciences,
IIT Kanpur
sudh@iitk.ac.in

1 Introduction

A motion event typically consists of Motion per se, a 'Figure' (an object which changes its location), a Ground (an object which serves as a reference point for the Figure's motion) and Manner (Talmy 2000b). Wide crosslinguistic differences are observed in encoding of these components in syntactic elements. In English, Manner is encoded in the verb (e.g. *The ball floated into the cave*) while in Spanish, it is encoded as adjunct to the main verb which encodes Motion and sometimes Path (e.g. *entrar en la cueva flotando* 'enter the cave floating'). Since the main verb encodes Motion and Manner in English, Path is encoded in satellites outside the verb. The question I address in this paper is what is encoded in the Kannada verbs – Manner or Path. While the linguistic analysis shows that both the kinds of encoding patterns are available, the elicited narratives show that speakers largely prefer to encode Path in the verbs rather than Manner.

2 Kannada – where does it fall on the typology?

Kannada belongs to the South Dravidian sub-branch of Dravidian family along with Tamil, Malayalam and Tulu among others. Nouns are inflected for cases and number; and verbs are inflected for tense, aspect, and agreement. The regular word order in Kannada is SOV, but in consonant with the extensive case marking and verbal agreement, the word order is relatively free and omission of both external and internal arguments is permitted.

Regarding encoding motion events, Kannada indeed allows some Manner verbs as the main verbs in some cases. With the Manner verbs, the Path is encoded in the satellites. Though Kannada lacks a system of morphemes dedicated to Path marking like the English spatial prepositions and particles, it does have a set of nouns indicating regions of space, and thus known as 'spatial nominals' and case inflections. They are used individually and in combination to encode various Paths: cf. *meele* 'on/above', locative case *–alli*, dative case *–ge*, ablative case *–inda, oLagininda* 'from the inside of/ through':

(1) banDe beTTadinda keLakke uruLitu
rock hill.abl downside.dat roll.pst.3sn
'The rock rolled down the hill'.

However, only translational motion verbs denoting "a motion forward along a horizontal straight line" (Talmy 2007: 81) can appear as the main verbs. The self-contained motion verbs denoting "a motion iteratively up and down along a vertical straight line" (Talmy 2007: 81) cannot; instead, they are encoded as an adjunct (as in 2):

(2) naanu kuNiyutta meTTilu iLide
I dance.pres.ptpl stairs descend.pst.1s
'I descended the stairs dancing'

Then does Kannada behave like Spanish? Kannada has a decent repertoire of Path verbs which can be classified into two main classes: (i) change of location (e.g. *daaTu* 'go-across', *haayu* 'go-by'); and (ii) posture change verbs (e.g. *malagu* 'lie down', *eeLu* 'get up'). In such cases, the Manner is encoded as an adjunct:

(3) Hemant hoLeyannu iiji daaTida
Hemant river.acc swim.pst.ptpl
cross.pst.3sm
'Hemant crossed the river swimming'

However, Path encoding verbs are fewer in number in comparison with a rich repertoire in Spanish or Korean. Also, the cases particularly the dative and the ablative are used extensively to indicate the Paths 'to' and 'from'.

So, the question is how do Kannada speakers encode motion events? Will they encode Path in the verbs or in satellites? Will they encode Manner in the verbs or as adjuncts? Or will they typically, use a 'neutral verb of motion' (such as *go, come*) (Slobin 2006) to describe the Figure's movement since encoding Manner in adjuncts is heavy on processing?

In order to investigate these issues, an elicitation study was conducted. The findings are reported in the next section.

3 Elicitation study

In the study, a total of 26 Kannada speakers (across five age groups i.e., from 4 year to 7 year and an adult control group) narrated two wordless picture stories viz., the Monkey story (Stringer 2005) and Alladin and the magical lamp (Stringer et. al 2011). During the elicitation, all kinds of references to Path or Manner were avoided in the prompts. The utterances were manually coded for the encoding of Path and Manner.

In the narratives, when the speakers encoded Path they did not have any preference; they encoded it in both verbs and satellites (see Figure 1). For example, a 4 year old child encodes Path in verb in an utterance - *mangaNNa meele hattitu* 'the monkey climbed-up/ ascended the top (of the hill)' – and in a satellite just a few minutes later - *mangaNNa marada oLage hooytu* 'the monkey went inside the tree'. The four age groups did not show a uniform behaviour in lexicalizing Path. The youngest group (4 year olds) preferred to encode Path in verbs (69.4%) than in satellites (30.7%). The other three groups (5-7 year old) encoded Path predominantly in satellites than in verbs. In the adult control group, however, encoding of Path in verb increased compared to these groups. A chi-square test showed that this difference across groups was significant (χ^2 = 33.75, df 4, p<0.001). These findings reflect the linguistic analysis where we observed that both the patterns are available in Kannada.

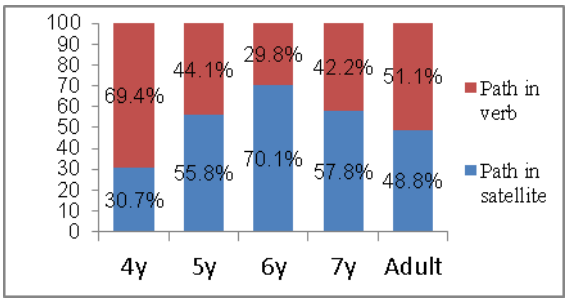


Fig1: Encoding Path in verbs and satellites

In the narratives, the percentage of utterances where Path-Manner were encoded in the same clause (i.e., Manner in the verb and Path in satellites) were few in number. A Chi-square test showed that there was no significant difference across groups: (χ^2=4.37, df 4, p=0.35). The participants were found to encode only-Path in motion events more often (4y: 55.3%; 5y: 49.1%; 6y: 50.9%; 7y: 55.6% and Adult control: 47.9%) than only Manner (4y: 13.2%; 5y: 20%; 6y: 9.8%; 7y: 2.2% and Adult control: 11.6%). Only-deictic verb utterances ranged from 9-24%. See Table 1 for details.

	4y	5y	6y	7y	Adult
Only-Path	55.3	49.1	50.9	55.6	47.9
Path and Manner	22.4	18.2	26.4	26.7	15.7
Only-Manner	13.2	20	9.8	2.2	11.6
Only-Deictic verb	9.2	12.7	12.9	15.6	24.8

Table1: Percentage of Semantic elements in Kannada motion events

The youngest group used only 4 Manner verbs and they were just 25% of total (verb + satellite) constructions. The highest number of Manner verbs was found in the adult speakers' narratives (5 in number). A more detailed analysis showed that in the Kannada narratives simple motion verbs plus

dative/ ablative cases (e.g. *mangaNNa marada oLage hoytu* 'monkey went inside the tree') constituted a large portion of satellite-framed constructions in all the groups. They show that even though Kannada has satellites and Manner verbs, (Manner verb + satellite) constructions are not as productive as those in English.

Acknowledgements

Many thanks to Dr.Shruti Sircar and Prof. R. Amritavalli for help with data collection and analysis.

References

Slobin, D. I. 2006. "What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition". In M. Hickmann & S. Robert (eds.), *Space in languages: Linguistic systems and cognitive categories* (pp.59-81). Amsterdam/ Philadelphia: John Benjamins.

Sudharshana, N. P. 2013. *Encoding spatial events in English as a second language: universal and language-specific features.* Unpublished PhD thesis. The English and Foreign Languages University. Hyderabad.

Talmy, L. 1985. Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description, vol. 3: Grammatical categories and the lexicon*. 1st edn. (pp.57-149). New York: Cambridge University Press.

Talmy, L. 2000a. *Toward a cognitive semantics Vol. 1: concept structuring systems*. Cambridge, MA: MIT Press.

Talmy, L. 2000b. *Toward a cognitive semantics Vol. 2: typology and process in concept structuring*. Cambridge, MA: MIT Press.

Talmy, L. 2007. Lexical typologies. In T. Shopen (ed.). *Language typology and syntactic description, Vol.3: Grammatical categories and the lexicon* (2nd edn) (pp. 66-168). New York: Cambridge University Press.

Formation of words in Chhatthare Limbu

Govinda Bahadur Tumbahang
Centre for Nepal and Asian Studies,
Tribhuvan University

Limbu is classified as Sino-Tibetan, Tibeto-Burman, Western Tibeto-Burman, Himalayan, Kiranti, Eastern language (Ethnologue 2012). Chhatthare Limbu is considered to be one of the four dialects of Limbu (Wiedert and Subba 1985, Van Driem 1987 and Kaila 2002). However, it is different from Panthare Limbu, the standard dialect of Limbu in vowel length, number of consonant phonemes, consonant clusters and morphemes. On the ground of common culture and ethnicity alone, it has been accepted as a dialect. But now, as Interim Constitution of Nepal 2007 has entitled every citizen to acquire primary education for their children in the mother tongue up to the primary level, children of the Chhatthare Limbu feel the need of their own mother tongue instead of the standard dialect of Limbu to get primary education because they cannot understand it. Children of other dialects of Limbu have no such problem since they understand it despite some variations in it.

Chhatthar area is located between the river Nubakhola in the east and Arun river in the west, Koshi river in the south and hill ridges in the north. Chhatthare Limbu is spoken in five Village Development Committees (VDCs) of Dhankuta district namely Tankhuwa, Teliya, Parewadin, Hattikharka and Murtidhunga and eight VDCs of Terhathum district namely Angdim, Hamarjung, Panchakanya Pokhari, Phakchamara, Okhre, Sudap, Dangapa, and Basantapur of eastern Nepal. Phedape Limbu is spoken across its eastern border, Panthare Limbu is spoken across its southern border, Lohorung Rai and Yamphu Rai are spoken across its northern border and Athpahariya, Bantawa and other Rai languages in and across the western border. Although some variations exist within it, they are intelligible to the Chhatthare Limbu speakers.

Royal Nepal Academy, a government institution has published a trilingual dictionary Limbu-English-Nepali dictionary (1961) and Nepal Academy republished it with considerable improvements in (2002). However, the dialect is purely Panthare in both publications. Recently, a grammar of Limbu (Limbu Vyakaran 2013) has been published by Nepal Academy, but it is the grammar of Panthare Limbu. Wiedert and Subba (1985) wrote a grammar of Panthare Limbu, Van Driem (1987) wrote a grammar of Phedappe Limbu and Mikhailovosky wrote a dictionary of Mewakhola Limbu dialect.

Until now, foreign and native linguists have not contributed to Chhatthare Limbu. Tumbahang (2007) and Tumbahang (2011) are the only works carried out on it. As germination and sequential cluster in the medial position of Chhatthare Limbu differ from other dialects of Limbu, they have impact upon stem alterations between vocalic and consonantal suffixes. So, these works have been focused on inflectional morphology.

However, the question as to how different words are derived in this language naturally emerges as a research question. This question is not answered adequately by the writers of the grammar of other dialects, too. Therefore, a study was conducted in order to discover the derivation process of new words from the existing ones employing questionnaire and focus group discussion methods, and has found derivational affixation, compounding and reduplication as the processes of formation of new words. Derivational morphemes comprise prefixes, circumfixes and suffixes. New words are derived by prefixation, suffixation and circumfixation. Compounding contains head modifier compounding, additive compounding and rhyming compounding. Reduplication consists of fully and partially reduplicated words.

tema (to come) is a verb with te- as a stem, and –ma is a suffix equivalant to English 'to'. It functions as a noun in a sentence. *phakma* (to fold) is a verb with *phak* as a stem. It is made adverb by adding the prefix ci to the stem *phak*. Consequently, it becomes an adverb *cibhak* which means "fully and at once". *lokma* (to run) is a verb comprising a stem *lok* and a suffix -ma. *kalokpa* is made by adding the prefix ka- and the suffix –pa to the stem surrounding it. These affixes constitute a circumfix, which conveys the meaning of the participial clause, "one which runs". Head modifier compound contains two nouns such as *phaksa* "pork", made of *phak* "pig" and sa "meat". The latter functions as a head and the former as a modifier. Additive compounding contains two nouns. The first noun bears the main meaning and the second conveys additional meaning, "and others". *tak* means "friend" and *luŋ* means "stone". When these two are combined as *takluŋ*, the latter noun loses its original meaning and conveys additive meaning "and others". Hence, takluŋ means "friends and others". Rhyming compounding is formed by repeating the phonemes of the rhyme in a syllable and forming a wrd like *tɔksɔk* "rice and the like" from tɔk "rice" . A word like *nubanuba* "good ones" is formed by fully reduplicating *nuba* "good" and a word like cujukpa "small ones" is formed by partially reduplicating *cukpa* "small".

References

Chemjong, I.S. 1961. *Limbu-Nepali-English dictionary*. Kathmandu: Royal Nepal Academy.

Eppele et.al.2012. Ethnologue: languages of Nepal. Kathmandu: LinSun and SIL International.

Kainla, Bairagi (edit). 2002. *Limbu- Nepali-English dictionary*. Kathmandu: Nepal Academy.

Mikhailovsky, B. 2002. *Limbu English dictionary of the Mewakhola dialect*. Kathmandu: Mandala Book point.

Tumbahang, G.B. 2007. *A descriptive grammar of Chhatthare Limbu*. A PhD. dissertation submitted to the Dean's office, Faculty of Humanities and Social Sciences, Tribhuvan University, Kathmandu.

Tumbahang, G.B. 2011. *A Linguistic study of Chhatthare Limbu*. Kathmandu: CNAS, Tribhuvan University, Kathmandu.

Van Driem, George.1987. A grammar of Limbu. Berlin, etc.: Mouton de Gruyter.

Weidert, Alfons and Bikram Subba.1985. *Concise Limbu grammar and dictionary*. Amsterdam: Lobster Publications.

A corpus based analysis of Hindi newspaper headlines representing specific socio-cultural practices

Vandana **Niladri Sekhar Dash**
Indian Statistical Institute, Kolkata
Vandana.mishra87@gmail.com ns_dash@yahoo.com
Jayshree Chakraborty
Indian Institute of Technology, Kharagpur
shree@hss.iitkgp.ernet.in

1 Introduction

This paper primarily aims to investigate the characteristic linguistic features of the newspaper headlines as reflected in some present day Hindi newspapers. It also wants to delve into the embedded implications of the headlines to understand the socio-cultural practices within a wider discourse of knowledge building and concept blending as represented in these headlines. Since the national-cum-social references include cognitive cues for knowledge representation, the people of the country can understand and recognize the reality around themselves through the constructs provided by the newspaper headlines. The socio-cultural constructs constitute "a whole complex of ambiguities and conventions' based on which the members of a community form their shared code and communicate with one another" (Moscovici 1984). The present paper wants to move in this direction to rip into the textual representation of Indian life and living, culture and customs, beliefs and thoughts through some Hindi newspaper headlines obtained from the corpus. Based on the analysis of texts taken from nearly 250 Hindi newspaper headlines, the paper tries to show how headlines showcase authentic representations of people's socio-cultural and national knowledge.

2 Relevance of the present study

The study of Newspaper headlines is an interesting field of research, as they occupy an important role by summarising the most important news of the day in a nutshell. In order to do so the headlines reflect varied linguistic structure in their choice of words or expressions which play a vital role in making these headlines more effective. Also, these Newspaper headlines are total discourse packages and their succinctness or grammatically incomplete status is not a barrier for the readers to have an overall understanding of the respective news discourses. This is because the words, expressions, punctuations they contain and also their syntax construct such context-specific meaning which socially, culturally and nationally aware members of the community can easily understand. So, newspaper headline provide a wide scope of research to observe and investigate any kind of linguistic phenomena that can be observed in this particular form of discourse.

3 Research Methodology

As the preliminary step to undertake the research is the collection of suitable data, so it was decided to collect about 250 headlines from the four prominent Hindi (Navbahrat Times, Amar Ujala, Dainik Jagran & Prabaht Khabar) dailies following a simple random sampling technique.

Genre	Amount
National news	60
International news	40
Business news	40
Sports news	30
Entertainment news	20
Regional news	15
Sci. & Tech. news	15
Life. & Culture news	10
Health news	10
Weather news	10
Total	250

The table 1 shows the amount of data collected from each domain which is in proportion to their amount in the whole corpus thus making the corpus representative enough for the present study.

4 Basic statistical information

Table 2 provides some basic statistical information about the dataset in quantitative terms of the sentences and words present in the sample corpus of 250 sentences.

Total no. Of sentences	250
Average length of each sentence	7.1
Total no. Of words (Tokens)	1679
No. Of Distinct Words	977
Type/Token Ratio	58.19
Average Word Length	4.11

Table 2: Basic Statistical analysis

Here, we find that the average sentence length in the newspaper headlines is 7.1 which are comparatively shorter than the length of Hindi sentences occurring in different kinds of other texts. The section also provides the total number of words in the corpus

(tokens), number of distinct occurrence of each word and also the average word length.

5 Typical linguistic features of the headlines

The headlines have a strong impact on the readers because certain linguistic features which make them particularly effectual. Most of the headlines occurring in the newspaper have few typical linguistic features. Some of the following linguistic features are appear in the Hindi news headlines:

(a) There are several varieties of sentence construction that appear in the Hindi newspaper headlines. The headlines usually are either compound or complex sentences on the basis of sentence construction and on the basis of meaning the news headline mainly contains interrogative, exclamatory, imperative and indicative sentences.
(b) The use of noun phrases is very prominent in the news headlines.
(c) The headlines are mostly in past/future tense with a finite verb; they are also in present tense but when the finite verb in absent. And the occurrence of finite verb results in present continuous tense in the headlines.
(d) The news headlines use gimmicks as in pun, alliteration, rhymes, personification etc. to create humour and ambiguity.
(e) The headlines also emphasize on the use of abbreviation. Most of these are the abbreviated forms of domain-specific terms present in the newspapers.

6 Brief discourse analysis of the headlines

The discourse analysis of the newspaper headlines attempts to show how the headlines reflect specific social practices and national ideologies which are mutually intelligible to people sharing the same cultural and social background. The understanding of grammar, morphology, semantics and phonology of a text does not always necessarily constitute understanding of the text (Kaplan 1990) but sometimes its meaning is reflected in the view of context in which it is being used. Some of these social and national beliefs represented through the headlines are:

a) आडवाणी ने चुनाव का शंख बजाया
Advani OBL. election POST. Conch blow.PERF.
'Advani blew conch of election'

In sentence (a) "शंख बजाया" (blew conch) represents a specific cultural practice signifying the commencement of any particular work or action.

(b) रेलमंत्री के आरोप पर माकपा गुस्से से 'लाल'
Rail minister POST Allegation POST. Maoist party anger POST. Red
Maoist 'red' with anger after rail minister accuses.

In sentence (b) the word 'लाल' (red) denotes anger and also denotes the identity of the particular political party as its flag is red in colour.

Also, the presence of code-mixed texts in the news headlines represents social cognition and awareness with innovative constructions which are not native to Hindi language but are used by the people in daily informal conversations.

(c) शॉपिंग: एडिक्शन या कुछ और?
Shopping addiction or something else
Shopping: addiction or something else?

In the sentence (c) the word 'शॉपिंग' (shopping) refers to the concept of buying/purchasing something which is now a generalized concept it used by people in their daily life and forms a part of their socio-cultural practice.

Thus the news headlines represent such socio-cultural beliefs and practices where the embedded meaning in these headlines can mostly be interpreted by people who are aware of these notions and culturally share the same knowledge.

7 Results & discussion

The findings from the present study reveal that the newspaper Headlines not only introduce or summarize the news articles but they also present the subjective attitude of the writer with their incredible linguistic choices to influence and shape the readers opinion who share the same cultural knowledge. This aspect might not be just limited to the "headlines" of the newspapers but its traces can also be seen in the other domains present in a newspaper. Hence, it can be concluded that newspaper as a form of discourse provides immense scope for further research focusing on textual analysis.

References

Alireza, B. and Samuel, M. 2012. Headlines in Newspaper Editorials: A Contrastive Study. *International Research Journal of Social Sciences* Vol. 1(3), 1-7; Available online at: www.isca.in

Bedřichová, L. 2005-2006. *Headlines and*

Subheadlines in Newspaper Reporting. Based on Discourse Analysis of Newspaper Articles in the Guardian. Available online at : http://is.muni.cz/th/39753/ff_m/Diplomova_prace.pdf.

Develotte, C. and E. Rechniewski. ????. *Discourse analysis of newspaper headlines: a methodological framework for research into national representations*. Available at http://wjfms.ncl.ac.uk/titles.htm

Fowler, R. 1991. *Language in the News: Discourse and Ideology in the Press* .Routledge. USA.

Tiono, N.I. 2003. An Analysis on Syntactic and Semantic factors Found in Newspapaer Headlines. k@ta . Vol 5, No 1; Available online at http://kata.petra.ac.id/index.php/ing/article/download/15497/15489

The emergence of the Ladakhi inferential and experiential markers out of a marker for admirativity: the case of *ḥdug*

Bettina Zeisler
Universität Tübingen
zeis@uni-tuebingen.de

The Western Tibetan markers for direct knowledge (experiential, immediate perception), on the one hand, and indirect knowledge (inference) and/ or admirativity, on the other, are apparently based on the same verb, Old Tibetan *ḥdug* with its lexical meaning 'stay'.

In almost all Ladakhi dialects, the marker for immediate visual perception takes the form *duk* (with the reduced forms *ruk* and *nuk* in certain phonetic environments and dialects).

Reduced forms of *duk* are attested in the Balti marker *suk* (used for personally perceived events, inferences, and surprises, Read 1934; Ghulam Hassan Lobsang 1995; Jones 2009), in the Shamskat Ladakhi past and perfect inferential marker *suk ~ sok* (own data), the Kenhat Ladakhi mirative marker *suk* (expressing surprise in Leh, Koshal 1979, and counterfactuals in Gya-Miru, own data), and further in the Kenhat past inferential marker *tok* (Koshal 1979), and in some Western Tibetan past or perfect inferential markers outside Ladakh: Kagate *tu ~ du*, Southern Mustang *tuk ~ ruk*, Lhomi *tuk*, and Jirel *duk* (Volkart 2000). The latter forms are quite apparently identical with the all-Ladakhi experiential marker *duk*. Similarly, the auxiliaries for generic knowledge and future inferences seem to be related: *(b)uk ~ (b)ok (< ba-ḥdug)* in western Shamskat (own data), *nok* in eastern Shamskat and Leh (Koshal 1979; own data), and *nak* in the Kenhat dialects further east (own data), cf. here the experiential auxiliaries *nuk* in Kyirong (Huber 2000) and Shigatse (Tournadre & Konchok Jiatso 2001) and *nok* in Sherpa (Volkart 2000).

There had been various attempts to derive the form *suk ~ sok*, but the most promising solution would be that *suk ~ sok* along with *tok* result from a contracted present perfect construction. This involves a morpheme *se (~ e)* in the Shamskat varieties of Lower Ladakh, a morpheme *ste (~ te)* in Leh, and a morpheme *te (~ de, re)* in the Kenhat varieties of Upper Ladakh. Hence we might have had a development *-seduk > *se'uk > suk* and *-(s)teduk > *te'uk or *t-duk > tuk*. The *o* vowel probably results from the lack of stress on the auxiliary. The past inferentials always imply some visual input concerning the resulting state of the event inferred.

They compete thus with the experiential present perfect with the auxiliary duk, cf. examples (1) and (2). Similarly, the generic and future inferentials are usually based on previous visual experience.

(1) *tiriŋ ane Tʃondrol ʂanʈe zap-tok.*
today aunt Chondrol very dress.up-INF.PA
/ zap-te-duk.
dress.up-PERF-EVID
'Aunt Chondrol apparently dressed up today. / has dressed up today (as I see from the result).' (Kenhat)

(2) *di phoŋ-khorpoa ramba ʂanʈe rgja-sok.*
this rock.PPOS:LOC ramba very grow-INF.PA
/ rgja-se-duk.
grow-PERF-EVID
'A lot of ramba (a kind of quitchgrass) has apparently grown / has grown (as I can see) around this rock.' (Shamskat)

Cross-linguistic data suggests that markers for indirect knowledge, particularly for hearsay must develop before markers for direct knowledge can evolve. In the Tibetic languages, however, the marker for immediate perception *ḥdug* appeared before the respective markers for hearsay, and in the non-western varieties it apparently also grammaticalised earlier than the markers for inference.

This anomaly can be traced back to an earlier admirative usage of the verb *ḥdug*. Even when *ḥdug* grammaticalised as a marker for immediate perception, it retained its connotation of non-commitment as opposed to authoritative knowledge expressed by the auxiliary *yod*.

It does not appear very intuitive that a position verb like 'stay' may develop evidential meanings. However, as a lexical verb, *ḥdug* 'stay' originally described a temporally extended, but not permanent position – in contrast to the existential linking verb *yod*, which described a general/ permanent existence or a more punctual location of an item at a certain place at a certain moment. These two temporal notions, and particularly the marked notion of preliminariness, could be exploited to differentiate between a preliminary truth, one the speaker would not fully commit to (*ḥdug*) and a general valid or authoritative truth (*yod*).

In its function as an existential linking verb and auxiliary, *ḥdug* started as a lexical or semi-grammaticalised marker of non-commitment or admirativity, indicating that a situation merely appeared to be as described, see example 0 – in this function, *ḥdug* could also be replaced by the lexical verb *snaŋ* 'to appear' or by the lexical verb *ḥdra* 'be like'. *ḥdug* may further indicate that the statement was based merely on an inference or reasoning, or that the speaker made a guess or prediction, example 0.

(3) *da bla.ma ḥdi⁺s-ni ḥbul.ba med-pa⁺r*
now lama this⁺ERG-Θ gift NG.have-NOM⁺LOC
gdams.ngag mi-gnaŋ-ba⁺r-ḥdug |
teaching NG1-grant-NOM⁺LOC-EVD.exist
'Now, it seems that this lama will not bestow the teachings [on me] without a gift.' (*Milaraspa rnamthar*)

(4) *«… da.lan ŋa⁺ḥi bu de thag.riŋ-žig-tu*
now.time I⁺GEN son that far-LQ-LOC
ma-soŋ-ba⁺r-ḥdug» *gsuŋ*
NG2-go.PA-NOM⁺LOC-EVD-exist say.PA
[The teacher asked his wife what Milaraspa was doing just now and she told him that Milaraspa has left in despair.] '[The teacher] said: «… By now, this my son, cannot have gone far.» (*Milaraspa rnamthar*)

In a second step, the notion of 'mere appearance' was extended to comprise also direct and immediate observations (by any person), example 0. In a last step, however, this latter usage became restricted to immediate observations made by the main speech act participant (that is, the speaker in assertions and the addressee in questions) concerning any other person and later also his or her own non-controlled accidental events (and states). The admirative and inferential function was retained for some time, and it can be argued that the evidential function of immediate perception was merely a more specific and more restricted subfunction of the notion of non-commitment or non-authoritative stance.

(5) *de⁺r rkos-pa-daŋ rdo⁺ḥi*
that⁺LOC dig.PA-NOM-COM stone⁺GEN
sgrom-nas chu ḥgro-ba gcig ḥdug-pa
box-ABL water go.PRS-NOM one exist.EVD-NOM
de bcad-te mtshon.cha⁺ḥi ḥkhor
that cut-NF sword⁺GEN wheel
thag-bcad-nas bltas-pa⁺s
rope-cut-ABL look.PA-NOM⁺INSTR
sgrom.chuŋ-žig-gi-steŋ-na yi.ge ḥdug-pa-la …
box.small-LQ-PPOS:LOC letter exist.EVD-NOM-LOC
bya-ba ḥdug-pa⁺s rgyal.po
say-NOM exist.EVD-NOM⁺INSTR king
khro-ste … bltas-pa⁺s | sgrom.chuŋ
be.angry-NF look.PA-NOM⁺INSTR box.small
de⁺ḥi-naŋ-na … nor.bu rin.po.che
that⁺GEN-inside-LOC jewel precious
ḥod ḥbar-ba bži ḥdug-pa⁺s
light burn.PRS-NOM four exist.EVD-NOM⁺INSTR
'[They] dug there, and (saw that) some water was coming out from a stone frame. They cut [it] off and so stopped the wheel with the swords [that protected the treasure], and when [they] had a look [at that place], ([they] saw that) there was an inscription on a small box, … and when it became apparent that it said: … , the king

became furious, and … when [he] looked [again], ([he] saw that) in side the small box, there were … four precious sparkling jewels' (*Biography of Chosrjedpal*)

The Tibetan systems of 'evidentiality' show great flexibility, allowing to scale up or scale down one's statement, which indicates that we do not deal with pure evidential systems, but that evidentiality has co-grammaticalised with speaker's attitude (or stance) and is thus dependent on pragmatic features.

Viewed from this perspective, both immediate perceptions and inferences belong to the domain of knowledge a speaker cannot (fully) vouch for – in contrast to knowledge that is based on long-standing experiences. It is, therefore, no longer surprising that the markers for immediate perceptions and for inferences have the same origin.

The Western Tibetan data further suggests that the phonetically reduced form *suk* reached the west in a first wave with a rather undifferentiated meaning of non-commitment, including the meaning of mere observation, while the reduced form *tuk* or *tok* must have reached the west somewhat later or on a separate route, without reaching the western-most areas of Baltistan and Lower Ladakh. *duk*, on the other hand, must have been introduced or re-interpreted as an auxiliary with the meaning of immediate perception in yet another later wave, which did not reach Baltistan and western Nubra. In the Ladakhi dialects, however, the introduction of *duk* as a marker of immediate perception lead to a more restricted reinterpretation of *suk* and *tok* as markers of indirect knowledge, that is, inference.

References

Ghulam Hassan Lobsang. 1995. Short sketch of Balti grammar. A Tibetan dialect spoken in Northern Pakistan. Bern: Universität Bern.

Jones, E. 2009. *Evidentiality and mirativity in Balti*. M.A. thesis, School of African and Oriental Studies, University of London.

Koshal, S. 1979. *Ladakhi Grammar*. Delhi etc.: Motilal Banarsidass.

Read, A.F.C. 1934. *Balti grammar*. London: The Royal Asiatic Society.

Tournadre, N. et Sangda Dorje. 1998. *Manuel de tibétain standard, langue et civilisation. Bodkyi spyiskad slobdeb*. Paris: Langues & Mondes/ L'Asiathèque.

Volkart, M. 2000. The meaning of the auxiliary morpheme 'dug in the aspect systems of some Central Tibetan dialects. In B. Bickel, ed., *Person and evidence in Himalayan languages*. Part I. *Linguistics of the Tibeto-Burman Area* 23(2): 127–153.