A very preliminary analysis of verb classification in the two Central Bunun dialects

Rik De Busser
National Chengchi University
rdbusser@nccu.edu.tw

1. Introduction

The Bunun language has five distinct dialects, which both native speakers and researchers consistently describe as mutually intelligible and which have been subdivided in a Northern, Central, and Southern branch. The two Central dialects, Takbanuaz (Jeng 1977) and Takivatan (De Busser 2009), are closely related historically and appear very similar in terms of their phonological structure, lexicon (Li 1988) and grammatical surface features. In plain language, to native speakers and students of the language they sound and behave more or less the same.

A reasonable assumption would then be that their grammars rely on identical (or at least very similar) morphosyntactic categories and that these categories are defined in terms of the same or at least similar formal and functional criteria. This talk will test this hypothesis through the comparison of verb subcategorization in Takbanuaz and Takivatan.

Based on data from narrative texts and directed elicitation, we will describe notable discrepancies in the surface morphology of verbs between the two dialects. We will investigate to what extent this leads to deviations in how the two dialects distinguish subclasses of verbs and how each of these classes is defined.

Sasse (1993) distinguishes functional, syntactic, pragmatic and morphosyntactic criteria for distinguishing word classes in a cross-linguistic context. The present discussion will focus on morpho-syntactic criteria involved, in particular the configurations of verbal affixes, such as voice markers and subclass markers, that are typically associated with verb classification.

What is presented below is a preliminary analysis based on data that is only partly processed.

2. Methodology

2.1. Data collection

Data for this study comes from two sources. The Takivatan Bunun data consists of an existing corpus of narrative texts and (to a much lesser extent) isolated examples that are the result of free elicitation or incidental recordings.

The Takbanuaz Bunun data was gathered during the present project and is the result of direct elicitation. Judgements about different verb forms and associated example sentences were collected through directed elicitation as follows:

- An ordered frequency list of verbal roots was created based on the Takivatan Bunun corpus and suitable candidates for elicitation were selected based on this list. This was used as a loose guideline for elicitations.
- Complex forms containing these roots, and other suitable verbal forms that came up during the elicitation sessions, were presented to the language consultant. An attempt was made to check a similar range of complex forms for each root. This was normally possible for voice forms and common prefixes, but difficult to attain for other forms. Forms that the consultant volunteered where also included.
- The consultant was asked (1) to judge whether the form was acceptable and (2) give a number of examples to illustrate its use.

An (incomplete) set for the root *tapha* ‘roast’ in Takibanuaz is shown below:

(1) ma-tapha titi
DYN-roast meat
Roast meat.

(2) ma-baliv tatapha?an
DYN-buy oven
Buy an oven

(3) ni asu ma-tapha
NEG 2S.TOP DYN-roast
Don't you want to roast something?

(4) na tapha-un a titi di
IRR roast-UV LNK meat here
Do you want to roast this meat here?

(5) * ta-tapha-an ludun titi
CV-roast-LV mountain meat

(6) in<in>han sak ludun di ma-tapha titi
<PST>inhabit 1S.TOP.AG mountain here DYN-roast meat
I roasted meat on this mountain here

(7) madia t<in>apha titi
many <PST>roast meat
There is a lot of meat that already finished cooking.

In the results below, we then compared the results about the acceptability of morphological verb alternations to the Takivatan corpus. This gives us some indication as to the similarities and differences between the two dialects, but this is not a symmetrical comparison. In the remainder of this project, I will attempt to gather data on Takivatan using the same elicitation methods as described above. Further research is necessary to expand the set of verb roots.

3. Preliminary results

The eventual aim of this research is to test (1) whether morphological verb alternations can reliably be used as indicators of verb subclassifications and (2) what the extent of variation is between the two closely related central dialects.

In a first step, we need to look at whether these two goals are feasible. Below are elicitation sets for two stative verb roots. Potential discrepancies between Takbanuaz (TBZ) and Takivatan (TVN) data are marked in bold. (Note that the Takbanuaz data indicates whether a form was accepted or not by the consultant, whereas the Takivatan data indicates whether a form can be attested in the corpus or not. This means that the absence of forms in Takivatan that are present in Takbanuaz are not necessarily meaningful. However, the absence of forms in Takbanuaz that are present in Takivatan are.)

Table 3.1 shows all elicited alternants of *sihal* 'good', a form which a very high frequency of use and a relatively large number of morphological alternants. Table 3.2 shows all elicited alternants of *daqvas* 'tall', a form with a low frequency and a moderate number of alternants.

A first observation that can be made is that both verbal roots allow for a large number of morphological alternants (18 and 13), but that accepted alternants do not necessarily correspond between *sihal* and *daqvas*, despite the fact that both are stative roots.

Second, there are much more attested forms for *sihal* (16) than for *daqvas* (1). This can be contributed to the absolute frequency of use, and is therefore not significant, but frequency effects like this might make a more systematic comparison difficult.

Third, first indications are that meaningful morphological differences between verb classes in the two dialects appear to be relatively few. However some of these differences warrant further investigation.

There is a lot of information in the two tables. We will here particularly focus on two phenomena: (1) the voice markers for UV and LV and (2) the prefix *tin-* 'SUDDEN'.

*Table 3.1. Elicitation set for **sihal** 'good' (35 forms / 18 acceptable / 16 attested)*

			TBZ	**TVN**
		Form	**Acceptable**	**Attested**
ROOT		sihal	yes	yes
VOICE	**AV**	masihal	yes	yes
	UV	sihalun	yes	yes
	LV	sihalan	**yes**	**no**
	PST	sinihal	no	no
	PST+UV	sinihalun	no	no
	PST+LV	sinihalan	no	no
INCH	**BASE**	minsihal	yes	yes
	BASE+LV			
	CAUS	pinsihal	**yes**	**no**
	CAUS+PST	pinisihal	yes	yes

	CAUS+UV	pinsihalun	yes	yes
	CAUS+LV			
	ASSOC	kinsihal	no	no
RECIP		pakasihal		yes
DYN	**AV**	masihalan	yes	yes
	AV+UV			
	AV+LV			
	AV+PST			
	CAUS	pasihal	**no**	**yes**
	CAUS+UV	pasihalun	possibly	yes
	CAUS+LV	pasihalan		no
	ASSOC	kasihal		no
	ASSOC+UV	kasihalun		yes
	ASSOC+LV	kasihalan		yes
STAT	**CAUS**	pisihal	yes	yes
	CAUS+UV	pisihalun	yes	yes
	CAUS+LV	pisihalan	no	no
	ASSOC	kisihal	no	no
	ASSOC+UV	kisihalun	**no**	**yes**
INSTR	**BASE**			
SUDDEN		tinsihal	**yes**	**no**
VARIOUS		palsihalav	**yes**	**no**
		palsisihalav	**yes**	**no**
		pinsihalav	**yes**	**no**
		ispinsihal	**yes**	**no**
		ispisihal	**?**	**yes**
		ispinsihalun	no	no
		ispalsihalav	no	no
		sinahalun	no	no
		tasihalan	**yes**	**no**

Let's first have a look at basic voice alternations. Both in Takbanuaz and Takivatan, stative verbs such as *sihal* 'good' can occur in non-actor voices (this has been previously observed, for instance in De Busser 2011). This indicates that voice marking can probably not be used as an indicator of verb subcategorization in the language. For Takbanuaz, we have a regular sets of voice alternants. The actor voice gets the stative prefix *ma-* and is unmarked by suffixes. The prefix is dropped in the undergoer and locative voices, which have the normal suffixes *-un* 'UV' and *-an* 'LV'.

(8) ma-sihal kaun-un
STAT-good eat-UV
It is very nice to eat

(9) sihal-un aipa saipuk-saipuk
good-UV DEM.S.DIST.VIS RED-take.care.of
(You/we) have to help him well

(10) sihal-an sak laihli di
good-LV 1S.TOP.AG car here
I really like this car here

Almost always, the form *sihalan* appears to refer to a positive affect of some sort. In most instances, it can be translated as ‘like’ or ‘feel for’.

(11) sihal-an sak babu di
good-LV 1S.TOP.AG pig here
I really like this pig here

This is compatible with the general tendency of verbs of emotion and perception to appear in the locative voice to foreground the theme of the action. (Non-canonical marking of arguments related to emotion is not uncommon cross-linguistically; see Aikhenvald, Dixon & Onishi 2001.) It is not clear at this point whether standard locative interpretations of LV forms of *sihal* are possible; so far, attempts to elicitation have been unsuccessful.

The form *sihalan* ‘good-LV’ is absent in Takivatan. This might have simply indicated a lacuna in the Takivatan corpus (after all, locative voices have a lower frequency than AV and UV), where it not for the existence of Takivatan locative voice forms which also include a prefix *ma-*, in in the example below:

(12) ma-sihal-an dalaq-un-a
MA-good-LV ground-UN-LNK
The land there was very good (Takivatan)

It is not entirely clear whether *ma-* should here be interpreted as a stative marker or has some other function. What is clear is that this construction also occurs with other stative verbs (see below).

*Table 3.2. Elicitation set for **daqvas** 'tall' (19 forms / 13 acceptable / 1 attested)*

			TBZ	TVN
		Form	**Acceptable**	**Attested**
ROOT		daqvas	**yes**	**no**
VOICE	**AV**	madaqvas	yes	yes
	UV	daqvasun	dubious	no
	LV	daqvasan	no	no
	PST			
INCH	**BASE**	mindaqvas	yes	no
	BASE+LV	mindaqvasan	no	no
	CAUS	pindaqvas	dubious	no
	CAUS+LV	pindaqvasan	yes	no
	ASSOC			
DYN	**AV**			
	AV+UV			
	AV+LV	madaqvasan	yes	?
	CAUS			
	ASSOC			
STAT	**CAUS**			

	CAUS+UV			
	CAUS+LV	pidaqvasan	yes	no
	ASSOC			
INSTR	**BASE**	isdaqvas	no	no
	CAUS			
SUDDEN	**BASE**	tindaqvas	yes	no
TIMESPAN	**BASE**	taldaqvas	yes	no
	CAUS	paldaqvas	no	no
VARIOUS		pindadaqvas	yes	no
		mundaqvas	no	no
		daldaqvasaŋ	yes	no
		ispindaqvas	yes	no
		istaldaqvas	no	no

In Takbanuaz, only the actor voice form of *daqvas* ‘tall’ appears to be accepted without reservations. The undergoer voice form can be used, but the consultant had considerably more reservations about its acceptability.

(13) ma-daqvas daiŋʔað-a Tiaŋ-un
STAT-high extremely-LNK PersName.M-UN
Tiang is especially tall

(14) daqvas-un makaiha
high-UV hang.over
Put it higher to hang over a line (e.g. of clothes).

A ‘naked’ locative voice form without prefix appears to be impossible for *daqvas*. However, a form with both a stative prefix *ma-* and LV *-an*, mirroring the Takivatan form *masihalan* above.

(15) maka-han ma-daqvas-an kusbai
PERL-go.to STAT-high-LV fly
It (e.g. the plane) flies over very high.

(16) ai maq a Huli tu bukðav daiŋʔað-a ma-daqvas-an
SURPR DEFIN LNK Puli ATTR plains extremely-LNK STAT-high-LV
As for the large plain of Puli, it is very high.

The two examples indicate that LV *-an* tends to get a more literal locative interpretation for the root *daqvas* than it did for *sihal*. This indicates that voice selection is observably influenced by semantic factors particular to individual verbal roots, which is a problem when we want to use them in verb subcategorization.

It is difficult to make any generalizations about the root *daqvas* in Takivatan, since only the form with a stative prefix *ma-* has been unambiguously attested.

(17) masmuav-in Bantalaŋ ma-daqvas
more-PRV Amis STAT-high
The Amis are very tall, … (Takivatan)

The above seems to suggest that the behavior of voice markers is not a very reliable indicator of verb subcategorization. Interestingly, a prefix that in our investigations of

Takbanuaz occurs fairly reliably on stative roots is *tin-*, which indicates that a state suddenly has come into being.

(18) tin-sihal-in nak tian
SUDDEN-good-PRV 1S.N belly
My belly is suddenly better.

(19) tin-daqvas-in-a uvaðʔað di
SUDDEN-high-PRV-LNK child here
This child has grown tall (suddenly).

The fact that these forms have not been attested in Takivatan might just be frequency effects. This opens up the interesting possibility that certain peripheral, low-frequency affixes might be better indicators of verb subcategorization than central affixes, such as voice markers. It is not clear at the moment what the theoretical implications would be, but practically this might be a very useful property to work with: peripheral prefixes tend to occur in simpler morphological constructions than central ones.

4. Discussion

4.1. Methodological limitations

The method described above is relatively efficient in collecting targeted information about morphological verb alternations, but during its implementation we encountered a number of issues.

First, our method is not objective and not completely systematic. The interviewer makes deliberate choices in selection alternations that they think are relevant to verb classification. Subsequent answers and follow-up questions depend often on thematic choices, unexpected linguistic discoveries, and often whim. It is almost impossible to make a consultant go through a fixed list of morphological alternations for a set of verb roots, and a fair amount of improvisation is needed.

It is also not clear what the exact criteria where that the consultant used to reject certain forms, beyond their general 'feel' for the language. For instance, it has been observed before that many Austronesian languages in Taiwan are very context-sensitive, and the absence of a suitable discourse context might have contributed to the rejection of certain forms (for instance, more unusual voice forms). We tried to compensate for this, and our consultant is linguistically aware.

This is especially the case given the large number of potential affixes that can attach to any verbal host. In reality, the combinations are restricted to various degrees, but this needs to be tested. This can be very cumbersome to the consultant with very restrictive verbal roots, with which most affix combinations are invalid. None of the elicited sets are therefore even nearly exhaustive. This would be difficult to attain, given the number of potential affixes and affix combinations. The method used here is useful for investigating a relatively small number of verb roots over an extended period of time. This is especially the case since this data will ideally need to be cross-checked with two or more consultants for each dialect.

Finally, as indicated above, the comparison between the two dialects is at present asymmetrical. I rely on elicited data for the Takbanuaz dialect and on textual data for

Takivatan. In the near future, I hope to remedy this by collecting elicited Takivatan data that mirrors the Takbanuaz forms.

4.2. Analysis

This paper is best seen as a pilot study that aims at investigating (1) whether morphological markers on verbs can be reliably used for verb subclassification and (2) how much morphological variation exists between verbal morphology in the two Central dialects of Bunun.

Even despite the limited data set, a number of preliminary conclusions can be drawn. In terms of verbal morphology, there appears to be relatively little difference between dialects, but certain differences might have a relatively large impact, especially when they concern voice. We discussed a possible difference between how LV *-an* operates in the two dialects. In addition, the voice options for the dynamic causative marker *pa-* appear to vary between Takbanuaz and Takivatan.

All in all, variation in voice markers seems to be at least partly determined by the semantics of individual roots. My impression is that this is also the case for certain other common valency-changing markers, such as the instrumental marker *is-* and its combinations with other affixes. This might indicate that central, high-frequency affixes are no good indicators of verbal categories.

Interestingly, there is some limited evidence that peripheral affixes like the sudden inchoative *tin-* might be used for verb subcategorization.

4.3. Future research

A lot of work remains to be done. The data set for Takbanuaz will need to be expanded to cover a larger number of forms of a larger number of verbal roots. Parallel elicitation will also need to happen for Takivatan.

The present analysis suggests that it might be worthwhile to spend more time on investigating relatively rare peripheral affixes in determining verbal categories. Given the complexity of verbal affixes and their combination, it might be necessary to investigate the role of transitivity, or more generally argument expression in the future.

More generally, an evaluation needs to happen how reasonable the present approach is for a large-scale classification of verbs. It might turn out that it is simply too labour-intensive to go beyond small-scale qualitative studies.

Bibliography

Aikhenvald, Alexandra Y., Robert M. W. Dixon & Masayuki Onishi (eds.). 2001. *Non-canonical Marking of Subjects and Objects* (Typological Studies in Language 46). Amsterdam: John Benjamins.

De Busser, Rik. 2009. *Towards a Grammar of Takivatan Bunun: Selected Topics*. Melbourne: La Trobe University PhD.

De Busser, Rik. 2011. Towards an analysis of argument alignment in Takivatan Bunun. *Studies in Language* 35(3). 523–555. doi:10.1075/sl.35.3.02deb.

Jeng, Heng-hsiung. 1977. *Topic and Focus in Bunun* (Institute of History and Philosophy Special Publication). Taipei: Academia Sinica.

Li, Paul Jen-Kuei. 1988. A comparative study of Bunun dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 59(2). 479–508.

Sasse, Hans-Jürgen. 1993. Syntactic Categories and Subcategories. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: Ein internationales Handbuch zeitgenössischer Forschung / An International Handbook of Contemporary Research, 1. Halbband / Volume 1* (Handbooks of Linguistics and Communication Science 9.1), 646–686. Berlin: Mouton de Gruyter.