

**Subgrouping in Tibeto-Burman:
Can an individual-identifying standard be developed?
How do we factor in the history of migrations and language contact?**

Randy J. LaPolla
City University of Hong Kong
ctrandy@cityu.edu.hk

Introduction

Two problems cloud our understanding of subgrouping in Tibeto-Burman. One is the lack of consistent and clear standards and principles for subgrouping. Subgrouping is often based on certain features that the languages are said to share, or on a few shared lexical items, or even on the fieldworker's intuitions, or on how remote speakers feel different languages are (the degree of mutual intelligibility). As discussed in Nichols 1996, in order to use the comparative method to establish genetic relatedness, it is necessary to find correspondences that statistically have an extremely low possibility of occurring as a matter of coincidence, that is, correspondences that have a one in a hundred thousand chance of occurring accidentally. In the Indo-European languages the morphological paradigms can serve this purpose. In monosyllabic morphology-poor languages achieving this standard is more difficult, but not impossible. What is needed is for cognate elements (words or morphological markers) to be organized into paradigmatic sets and applied rigorously to determine relatedness. That is, the *sets* are reconstructed and used as individual-identifying evidence of relatedness, because the particular combination of cognate elements as an internally-structured set would give us the level of statistical significance we need. We will attempt to create an example of such a set using the person marking and reflexive/middle marking and certain other distinctive features found in a subset of TB languages that can be taken together to identify the individual protolanguage from which those languages descended, and so identify that subset of languages within TB as a valid genetic grouping.

The second problem was raised in a recent study of the migrations of the Sino-Tibetan speaking people (LaPolla in press). If we look at the various proposals for subgrouping in Tibeto-Burman, almost all include geographic designations such as "Western Himalayan", "Eastern Himalayan", etc. The recent debate about the use of the term "Kamarupan" (LTBA 22.2) hinges partly on the validity of using geographic terms. These designations assume that either the languages involved have always been at that location or that all the languages

developed from a single ancestor which migrated to that location at some time in the past. Yet we know that there have been waves of migration, particularly in certain regions, and so there is a suspicion that the languages were not originally closely related, but have come to seem similar because of long-term contact. Even for groupings that are not given geographic names, geographic factors seem to have played a role in determining which languages are seen as part of that grouping. For example, the split between Bodic (given the definition I will propose) and the rest of TB was due to an early migration west and then south throughout Tibet and down into Nepal, Bhutan, and India. But some of the languages now classified as Bodic, such as the Kiranti languages and the Western Himalayan languages of Uttar Pradesh, show individual-identifying correspondences with languages that clearly did not come down with the Bodic migration, such as Rawang. The only possible conclusion is that those so-called Bodic languages did not come down from the north, but came across from the (south-)east, and show similarities with the true Bodic languages because of contact rather than genetic inheritance.

Type-Identifying vs. Individual-Identifying Evidence

Nichols 1996 is a discussion of what types of evidence we need to show relatedness between two languages. Nichols argues that the evidence that has been used in the history of Indo-European linguistics for showing relatedness is not individual word correspondences, but 'whole systems or subsystems with a good deal of internal paradigmaticity, ideally multiple paradigmaticity, and involving not only categories but particular shared markers for them' (p. 48). The reason for this is that there is very little likelihood of these paradigms appearing in different languages purely by chance, and so that paradigm can be said to have developed only once, and therefore any languages that share that paradigm must have developed out of the single language in which that paradigm developed. That is, the paradigm identifies a unique individual protolanguage. This type of evidence Nichols calls 'individual-identifying' evidence, and she opposes it to 'type-identifying' evidence, features such as configurationality, a particular word order, or the simple presence of a particular category such as gender or ergativity, evidence which may identify a type of language, but not a unique individual proto-language. Nichols establishes the statistical threshold for statistically significant individual-identifying evidence at one out of a hundred thousand. That is, the probability of a particular set of evidence appearing by chance in two or more languages should be less than one in a hundred thousand. This level is achieved by most of the evidence used for relatedness in Indo-European linguistics. Nichols (pp. 49-52) gives four

examples of such evidence, two grammatical and two lexical: the paradigm of *good* and *better* in English and its sister languages, the paradigm of gender marking in IE, the word for 'widow', and the structured system of numbers. The word for 'widow' is individual identifying because the chance of a language having four consonants reflecting the four PIE consonants **w*, **y*, **dh*, and **w* in that order (assuming each language has 20 consonants, a world-wide average) is $.05 \times .05 \times .05 \times .05 = .000000625$, less than one in a hundred thousand. In Sino-Tibetan we cannot achieve this level of statistical significance with single words, and there is also the factor of borrowing, so we must depend on paradigmatic sets or words and grammatical forms. It is paradigmaticity in particular that helps us reach the individual-identifying threshold, as the probability for the set as a whole is determined by multiplying the probabilities of the individual forms and categories by each other. For example, the paradigm of *good* and *better* achieves a probability of occurrence of 0.000000125, less than one in a million (that is, each word has a probability of .0005, assuming 20 consonants and five vowels per language— $0.05 \times 0.2 \times 0.05$ —but then because these two words form a paradigm, these two probabilities are multiplied together and by the probability of the two forms being a paradigm: $0.0005 \times 0.0005 \times 0.5$). In some cases structured sets of vocabulary items can achieve the level of individual-identifying significance, such as the set of basic numbers. The probability of having a set of numerals from 'one' to 'five' where the first two consonants form the pattern given in (1) is 0.98×10^{-13} (Nichols 1996:52).

(1)	1	y, n	$0.05 \times 0.05 = 0.0025$
	2	d, w	$0.05 \times 0.05 = 0.0025$
	3	t, r	$0.05 \times 0.05 = 0.0025$
	4	k ^w , t	$0.05 \times 0.05 = 0.0025$
	5	p, n	$0.05 \times 0.05 = 0.0025$
			$(0.0025)^5 = 0.000000000000098$

To use sets of lexical items as individual-identifying evidence, the set must be paradigmatically structured, and the entire set has to be attested in each language. Without this lexical comparison cannot be used as evidence of relatedness among isolating languages:

[A]ny claim of genetic relatedness among isolating languages that relies simply on lexical comparison—without (tonal or other) arbitrary lexical classification and without paradigmatic lexical sets attested as whole sets in

each language—probably cannot be regarded as individual-identifying and thus as consistent with the comparative method, no matter how numerous the compared lexemes. (Nichols 1996:64)

Therefore, if we want to find solid individual-identifying evidence in Sino-Tibetan, we should look for paradigmatic sets, or create sets of features. In Sino-Tibetan, many of the basic numerals have only a single consonant, and so the probability would be somewhat lower than the set of IE numerals, and there is also the problem that a set of numbers like this is precisely something that could be borrowed as a set. Therefore we would need to use the numbers together with other features to create a set that as a set is individual-identifying. For example, we might be able to use the numerals plus the voiced(or unaspirated) vs. voiceless(or aspirated) opposition in initial consonants (as well as cognacy) marking transitivity or simplex vs. causative, the **s-* causative suffix, the **-t* transitivizing suffix, and the **-n* formative suffix, if a survey of the languages showed each of these features did appear in all assumed major groups as a set of features that could identify with certain whether a language (or language group) is or is not Sino-Tibetan. I have not yet done the survey yet, but plan to after resolving some issues related to lower level groupings.

Nichols' main goal in her paper is to set up a clear criterion for determining if some feature is or is not useful in determining the relatedness of languages, and gives examples to show how the evidence garnered to support the Nostratic hypothesis does not reach the standard of individual-identifying evidence. One question that comes up in the application of this criterion, given a piece of evidence (such as a grammatical sub-system) that reaches the individual-identifying threshold and a set of languages that are assumed to be related, is 'In how many daughter languages or branches must the system be attested in order for it to confirm genetic relatedness?' (Nichols 1996:60). Nichols' answer to this question is

If it is the sole individual-identifying evidence for relatedness, then obviously it must be firmly attested in EVERY daughter branch (at the highest level at which daughter branches can be identified and reconstructed) . . . If it is not the sole evidence but simply one of several pieces, then its distribution among the daughter branches can be evaluated as a binomial distribution: ten out of twelve or nine out of ten is significant, six out of ten or three out of four is not, etc. . . . Since most deep protolanguages will have only a few first- and second-order surviving daughter branches (Nichols

1990), in practice this means that unanimous firm attestation among the daughter branches must be required of each piece of individual-identifying evidence.' (p. 60-61, emphasis in original)

In this passage Nichols is talking about the level of attestation required of a piece of individual-identifying evidence for it to be considered evidence of genetic relatedness. That is, if we find a particular individual-identifying set of features in all the branches of a recognized family, then that set can be used as a test for membership in the family. There is a second use, though, to which we can put this criterion. We can apply this same level of statistical significance in determining whether some feature should be reconstructed to the deepest level proto-language, or to some lower level of the family tree. That is, if we want to reconstruct some feature to Proto-Tibeto-Burman, given our minimum statistical threshold of significance (that of a binomial distribution) we can easily decide whether the feature should or should not be reconstructed to the PTB (or Proto-Sino-Tibetan) level. For example, person marking in TB only appears in about half of the branches of TB, and so does not reach the level of statistical significance required to be reconstructable to PTB. If it is to be reconstructed at all, it would have to be at some lower node in the tree, the daughter branches of which would all have to attest a cognate pattern of person marking. Turning this around, we can then use this piece (or set) of individual-identifying evidence to identify a particular subgroup within TB.

Tibeto-Burman

If we take the system of verbal suffixes related to the person-marking system of Proto-Dulong-Rawang as an example, we can calculate the odds that another language will have the same system as follows. The Proto-Dulong-Rawang system has a 1sg suffix **-ŋ*, a dual marking suffix **si*, a 1pl suffix **-i*, and a 2pl suffix **-n*. Assuming the other language has a person-marking system,¹ the probability of finding exactly a velar nasal suffix for 1sg marking and **-si* as a suffix for dual marking is $0.05 \times 0.5 \times (0.05 \times .2) = 0.00025$, or two in ten thousand. Add to this the 2pl suffix, and we bring the probability down to 0.0000008975 ($0.05 \times (0.05 \times .2) \times 0.05 \times 0.0359^2$), or roughly nine in ten

¹This is actually a probability that should also be factored into the overall probability, but as we do not have a representative number of languages with which to calculate it, we will have to simply calculate assuming we are dealing only with languages that have person marking.

²This is the chance of having the three forms code 1sg, 2sg, and dual, respectively ($0.33 \times 0.33 \times 0.33$). When we add in the 1pl marker below, the

million. Adding the 1pl suffix brings this down to 0.000 000 0195 (0.05 x (0.05 x .2) x 0.05 x 0.2 x 0.0039), or one in a hundred million. Even assuming a somewhat higher probability because we assume that the PTB 1sg pronoun was **ŋa* and the 2sg pronoun was **na(ŋ)*, and that the person marking would derive from the free pronouns, this is still well within the individual-identifying range. That is to say, the odds of finding an unrelated language (and we should say "non-contiguous language unrelated language" to minimize the possibility of borrowing) with this system are less than a million to one. One language that has been suggested but never proven to be related to Dulong-Rawang is rGyalrong (e.g. Chang & Chang 1978; Thurgood 1984). If we compare the Proto-rGyalrong agreement, as reconstructed by Qu (1983) with that of Proto-Dulong-Rawang, we find that there is a close, but not exact match.

	1sg	2pl	dual	1pl
Proto-rGyalrong	*-ŋ	*-ñ	*-tsh	*-i
Proto-D-R	*-ŋ	*-n	*-si	*-i

The odds that we could get this sort of match-up in two non-contiguous languages by chance are still well within the individual-identifying range, and so we can be certain that Dulong-Rawang and rGyalrong are related at some level.³ But this is only a single piece of evidence. To be really certain of relatedness, we should add other, unrelated features to the set. One feature we can add to the set is the Dulong-Rawang reflexive/middle suffix **si* (distinct from the dual suffix), bringing the overall probability for having both the person marking paradigm discussed above and the reflexive suffix **-si* down to 0.000000000 016, or one in a hundred billion (0.05 x (0.05 x .2) x 0.05 x 0.2 x (0.05 x 0.2) x 0.00032).

chances of the four forms coding exactly 1sg, 1pl, 2sg, and dual, goes down to 0.0039).

³A side note on the Qiangic languages: I had at first hoped to show that the set of directional prefixes could be used as a major piece of individual-identifying evidence for the Qiangic languages, but the pattern of correspondences does not lend itself to reconstruction. I also tried to see if Qiang and other languages in the group could be shown to be related to rGyalrong based on the person marking, but could not find a sufficient pattern of correspondence. Much more work must be done on this group before anything conclusive can be said about it.

We know that the majority of TB languages do not manifest this system, so it cannot be used as individual-identifying evidence for membership in the Tibeto-Burman family, but we can use it as very strong evidence of genetic relationship at some lower level. What I would like to suggest is that non-contiguous languages that can be shown to manifest this system must be considered a single genetic grouping distinct from groups that do not manifest it, because the odds of them sharing such a system by chance are infinitesimal. rGyalrong does not manifest the reflexive suffix **-si*, but there are languages that manifest both the person marking system and the reflexive suffix, such as the Kiranti languages and the Western Himalayan TB languages of Uttar Pradesh (Darma, Chaudangsi, Byansi, Rongpo, Raji) and Himachal Pradesh (Kinnauri) in Northeastern India.

In Dulong-Rawang (data from my own fieldwork), the pattern is relatively straightforward, except for the form of the 2pl marker, where we have *-n*, *-juŋ* and *-nuŋ*. As *-juŋ* is a pronominal plural marker (*əŋ* '3sg pronoun'; *əŋ-jùŋ* '3pl pronoun'), and *-n* occurs by itself in some dialects, it seems the form *-nuŋ* may be a fusion of an original 2nd person marker and a plural marker. For this reason I have reconstructed **-n* for the 2pl marker.

	1sg	1pl	2pl	dual	refl
Rawang (Matwang)	-ŋ	-i	-nuŋ	-ɕi	-ɕì
Rawang (Daru)	-ŋ	-i	-nuŋ	-si	-sì
Dulong (1st Township)	-ŋ	long vowel	-juŋ	-ɕu	-ɕu
Dulong (3rd Township)	-ŋ	-i	-n	-ɕu	-ɕu
Dulong (4th Township)	-ŋ	-i	-n	-ɕu	-ɕu
Nujiang dialect	-ŋ	--	-ŋuŋ	-ɕu	-ɕu
Proto-D-R	*-ŋ	*-i	*-n	*-si	*-sì

Ebert, in press, gives the following description of the Kiranti verb:

The Kiranti finite verb is characterized by a complex system of person and number affixes. Some widespread Tibeto-Burman suffixes are well preserved in Kiranti: *-ŋa* 1st singular, *-na* 2nd singular, *-i* 1st and 2nd plural (the latter also *-nin*), *-u* 3rd patient, *-ci*, *-si* dual. SE languages also have prefixes with parallels in other TB languages, like *kɛ-*, *ta-*, *a-* for 2nd person, *I-* for inverse, *i-*, *a-* for 'marked scenarios' (inverse configurations and 2nd person participation).

A reconstruction for Proto-Kiranti is given in van Driem 1990. I add the example of Dumi (van Driem 1993), to show the pattern in an attested language.

	1sg	1pl	2pl	dual	refl
Proto-Kiranti	*-ŋ	*-i	*-ni	*-ci	*-nsi
(e.g. Dumi)	-ŋ	-i(inclusive)	-ini	-si	-(n)si

The last group is the Western Himalayan languages. The data are from S.R. Sharma 1994, in press, ShreeKrishan in press, and D.D. Sharma 1988 (see p. 150 on the refl/intransitivizer; 2 sg -n)

Kinnauri	-g	-č	-č/-ñ	-č/-ñ	-si
Rongpo	-ŋ	-ni	-ni	--	-s
Byangsi	-Ø(yε)	nye	-ni	-ci	-ci
Darma	-Ø(yε)	--	-ni	--	-ci
Chaudangsi	-Ø(yε)	--	-ni	--	-ci
Raji	-Ø	-i (pl)	-i (pl)	--	--
Manchad	-g	-ñi	-ñi	-shi	--
Tinan	-g	-ñi	-ñi	-shi	--
PWH	*-g/ŋ	*ni	*-ni	*-si	*si

The forms given here are for present tense agreement. Quite a number of you are probably thinking to yourselves "Why didn't he include the 2sg *-n suffix? the answer is that we started with Dulong-Rawang, and Dulong-Rawang does not manifest the *-n suffix for 2sg, only 2pl. The Kiranti languages and the Western Himalayan languages do show evidence for a proto 2sg suffix *-n.

Proto-D-R	*-ŋ	*-i	*-n	*-si	*-si
Proto-Kiranti	*-ŋ	*-i	*-ni	*-ci	*-nsi
Proto-W. Himalayan	*-g/ŋ	*ni	*-ni	*-si	*si

The odds of achieving this close of a correspondence between unrelated non-contiguous languages is almost infinitesimal, but there is other evidence that we can add to this, to make the relationship even more clear. One feature that is shared among the three groups (although not attested in every language) is having a cognate word that means both 'sit' and 'live, dwell, (come to) be at', e.g. Rongpo (W. Him.) *huŋ* 'to sit, to live, to be'; Belhare (Kiranti) *yuŋ* 'sit; come to be at a place'; Dulong *rùŋ* 'sit; live in a place'.

A feature that might show a closer relationship between Kiranti and Dulong-Rawang than either has with the Western Himalayan languages, is the prefix that marks a situation where one of the arguments is a speech act participant, but the actor is not the speaker (in the quote from Ebert above, she characterizes it as being 'for "marked scenarios" (inverse configurations and 2nd person participation)'), e.g. Rawang *-e*, Dulong *nə- ~ -ε-*; Khaling *-i*, Dumi *-a*. We don't find clear cognacy of the vowel used for this marking, but its use is idiosyncratic enough that the likelihood of parallel development is rather low, and so WHEN USED WITH OTHER FEATURES AS PART OF A SET this feature can help us to identify a single parent language.

How do we explain these findings? I believe we must turn to the migrations as an explanation. As I discussed in LaPolla in press, It seems there were two major routes of migration for the ancient "Qiang" or TB speakers: west into Tibet and then south, and Southwesterly down along the eastern edge of the Tibetan plateau and across Burma into Northeast India and then up into Nepal. This latter route is the one I believe the ancestors of the Dulong-Rawang, Kiranti, and Western Himalayan peoples went. The point of all of this is that we need to keep in mind the fact that there has been wave after wave of migration, and not assume that all people in a particular area are necessarily related, even if they appear similar. Particularly for languages in known paths of migration, we must find clear evidence of relatedness.

References

- Chang, Kun & Betty Shefts Chang. 1975. Gyarong historical phonology. *Academia Sinica: Bulletin of the Institute of History and Philology* 46:391-524.
- Driem, George van. 1990. An exploration of Proto-Kiranti verbal morphology. *Acta Linguistica Hafniensia* 22.2:27-48.
- Driem, George van. 1993. *A grammar of Dumi*. Berlin: Mouton de Gruyter.
- Ebert, Karen H. 1987. Grammatical marking of speech act participants in Tibeto-Burman. *Journal of Pragmatics* 11.4: 473-482.
- Ebert, Karen H. 1990. On the evidence for the relationship Kiranti-Rung. *Linguistics of the Tibeto-Burman Area* 13.1:57-78.
- Ebert, Karen H. 1994. *The structure of the Kiranti languages* (Arbeiten des Seminar für Allgemeine Sprachwissenschaft Nr. 13). Zürich: Universität Zürich Seminar für Allgemeine Sprachwissenschaft.
- Ebert, Karen H. in press. The Kiranti languages. *The Sino-Tibetan languages*, ed. by R. J. LaPolla & G. Thurgood. London: Curzon Press.
- LaPolla, Randy J. 1992. 'On the dating and nature of verb agreement in Tibeto Burman', *BSOAS* 55.298-315.
- LaPolla, Randy J. 1994. 'Parallel grammaticalizations in Tibeto-Burman: Evidence of Sapir's 'drift'', *LTBA* 17.61-80.

- LaPolla, Randy J. 1995a. Reflexive and middle marking in Dulong/Rawang. Paper presented to the 28th International Conference on Sino-Tibetan Languages and Linguistics, Charlottesville, VA, October 6-9.
- LaPolla, Randy J. 1995b. On the utility of the concepts of markedness and prototypes in understanding the development of morphological systems. *Bulletin of the Institute of History and Philology* 66.4.
- LaPolla, Randy J. in press. The role of migration and language contact in the development of the Sino-Tibetan language family. *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*, ed. by R. M. W. Dixon & A. Y. Aikhenvald. New York & Oxford: Oxford University Press.
- Nichols, Johanna. 1990. Linguistic diversity and the settlement of the New World. *Language* 66:475-521.
- Nichols, Johanna. 1996. The comparative method as heuristic. *The comparative method reviewed*, ed. by M. Durie & M. Ross. New York & Oxford: Oxford University Press.
- Poffenberger, Mark. 1980. *Patterns of change in the Nepal Himalaya*. Delhi: The Macmillan Co. of India.
- Sharma, Devidatta D. 1988. *A descriptive Grammar of Kinnauri*. Delhi: Mittal Publications.
- Sharma, Suhnu Ram. 1994. Pronouns and agreement in West Himalayan Tibeto-Burman languages: A comparative statement. Paper presented at the 27th International Conference on Sino-Tibetan Languages and Linguistics, Paris, Oct. 12-16, 1994.
- Sharma, S. R. in press. 'A study on the Tibeto-Burman languages of Uttar Pradesh', in *The Tibeto-Burman languages of Uttar Pradesh*, edited by Randy J. LaPolla (Vol. II of *A Linguistic Approach to Zhangzhung and Related Languages in the Indian Himalayas*). Osaka: National Museum of Ethnology.
- ShreeKrishan. in press, a. 'Chaudangsi', in *The Tibeto-Burman languages of Uttar Pradesh*, edited by Randy J. LaPolla (Vol. II of *A Linguistic Approach to Zhangzhung and Related Languages in the Indian Himalayas*). Osaka: National Museum of Ethnology.
- ShreeKrishan. in press, b. 'Raji', in *The Tibeto-Burman languages of Uttar Pradesh*, edited by Randy J. LaPolla (Vol. II of *A Linguistic Approach to Zhangzhung and Related Languages in the Indian Himalayas*). Osaka: National Museum of Ethnology.
- Thurgood, Graham. 1985. Pronouns, verb agreement systems, and the subgrouping of Tibeto-Burman. *Linguistics of the Sino-Tibetan area: The state of the art*, ed. by Graham Thurgood, James A. Matisoff & David Bradley, 376-400. Canberra: Pacific Linguistics (C-87).
- Thurgood, Graham. 1988. The 'Rung' languages: notes on their proto-morphosyntax and subgrouping. *Acta Orientalia* 46: 79-99.
- Toba, Sueyoshi. 1988. The pronominal affixation system in Khaling. *Prosodic analysis and Asian linguistics: to honour R. K. Sprigg*, ed. by David Bradley, Eugénie J.A. Henderson and Martine Mazaudon, 201-204. Canberra: Pacific Linguistics C-104.