Comments on Methodology and Evidence
in Sino-Tibetan Comparative Linguistics

Randy J. LaPolla

La Trobe University

This paper presents epistemological and methodological problems found in work on the subgrouping of Sino-Tibetan languages and the reconstruction of features of the languages. A key problem is the lack of an accepted standard for judging this work, one that can stand up to statistical evaluation. An alternative methodology that involves using fixed sets of features to give us the statistical probability of common origin is suggested.

Key words: Sino-Tibetan, historical comparative method, statistical analysis, reconstruction methodology

1. Preamble

The present paper was initially prepared as a commentary on Guillaume Jacques' paper, "Sino-Tibetan morphology from a Rgyalrong perspective", presented at the International Symposium on Sino-Tibetan Comparative Studies in the 21st Century, Institute of Linguistics, Academia Sinica, June 24-25, 2010. Guillaume visited with us at the Research Centre for Linguistic Typology for several months prior to the Symposium, and so Guillaume and I had the chance to discuss the issues he raised in that paper at length, and in some cases he has accepted criticisms of things I thought were problematic in his analysis, but in other cases, including the whole way of writing such a paper on the basis of so much conjecture and assumption, we have agreed to disagree. Because he has now split that original paper into three separate papers, and because the difference in approach is more significant than the details of individual issues, in my comments here I would like to just discuss the issue of approaches rather than the details of his paper. I shall break it down into different types of problems related to approach.

2. Epistemological problem

The first problem we encounter is an epistemological one, "Teeter's Law" (Watkins 1976:310): "The language of the family you know best always turns out to be

the most archaic". This is similar to the Chinese expression "Blind men feeling an elephant" (瞎子摸象): one interprets the nature of the whole based on exposure to only one part of the whole. But we should not think of this simply in a negative way; we need to understand that this is a natural part of human cognition, and it explains much about how we understand the world. Underlying it is a general principle of interpretation: We understand (make inferences about) something by relating it to what we know already. When making inferences, just as what we already know will influence how we understand something, the initial inferences of that interpretive process will influence the rest of the process; that is, it will influence the construction of the context of interpretation.

Much of our cognition involves abductive inference, a kind of hypothesis creation, basically trying to "make sense" of something we have observed by creating a context in our minds (a context of interpretation) in which whatever we have observed will seem relevant and consistent with what we know already. This is the sort of inference that underlies religion, science, and communication (see LaPolla 2003b for discussion). When making such inferences, the initial contact with whatever it is we are trying to understand influences the initial construction of the context of interpretation, and so influences the understanding of whatever comes after it. So not only will we have a different understanding of the elephant if we only feel the tail as opposed to the body, we will also have a different understanding even if we then move on from one part and feel the entire elephant, as having started with the tail, for example, will influence how we understand the rest of the elephant, because it will be our reference point in understanding the rest of what we feel. This is why the Theme in the Theme-Rheme structure of the clause is important: as the first element encountered in the clause, it influences the interpretation of the rest of the clause (Halliday 1994: Ch.3). So it is natural that, when working on a language family, having a greater and previous knowledge of one of the languages will necessarily influence how one thinks about the family as a whole. We see this very clearly in Guillaume's reconstructions that make Proto-Sino-Tibetan look like modern Rgyalrong, and also in George van Driem's reconstructions that make Proto-Tibeto-Burman look similar to modern Kiranti languages (e.g. van Driem 1993). They have each taken one part of the elephant as representing the whole elephant. It is particularly telling that they chose different parts of the elephant to start with, Rgyalrong and Kiranti, respectively, and so came to different conclusions.[1] In Guillaume's case the assumption that Proto-Sino-Tibetan must have been morphologically complex, even though all of the oldest written languages are not, is due simply to his assumption that

---

[1] There are of course similarities in their reconstructions because, as I have argued separately (LaPolla 2000), the Kiranti and Rgyalrong groups are closely related, but still the differences in the reconstructions due to the different starting points are clear.

Rgyalrong morphology is the most archaic, an assumption that is due simply to his having worked extensively on Rgyalrong.

This sort of assumption brings us to a second part of the epistemological issue mentioned above: the *a priori* assumptions we bring to the problem will influence what we see and how we analyse it.

As mentioned above, what is already in our minds, that is, what we know, what we have experienced, influences how we understand new things and experiences, as the things "that we know" become assumptions used as part of the context of interpretation in understanding the new phenomenon. They can even influence our perception of the phenomenon. The different assumptions we bring to the problem of reconstruction and genetic relations influence our understanding of the whole issue. For example, if we assume that Sino-Tibetan must have been morphologically complex, that will influence our understanding of new data and the conclusions we make, compared to if we work inductively, not assuming anything *a priori*. Another example is that the assumption of a particular Chinese or Tangut reconstruction system will influence what words we see as cognate. In his 1979 paper, Mei Tsu-Lin thought that the Tibetan cognate for the Chinese word *yè* (夜) 'night' should have been *zla* 'moon', based on the assumption that the Chinese form should be reconstructed with an open final (based on Li 1971). Guillaume, on the other hand, assumes the Tibetan cognate for this is Tibetan *zhag* '24 hour period' because he assumes a Chinese reconstruction with a velar stop (based on Baxter 1992).[2] A third example is the key factor underlying the disagreement about the question of the reconstructability of the person marking found in some Tibeto-Burman languages to the Proto-Tibeto-Burman stage: whether we assume that a form that is a transparent grammaticalization should be reconstructed to the proto-language or not. In my work I assume that if a morphological form is a transparent grammaticalization, it should be a recent innovation, and I assume in terms of methodology that one only reconstructs those items of morphology for which we cannot see any obvious source in grammaticalization, and so I do not think we should reconstruct person marking to Proto-Tibeto-Burman (see LaPolla 1992 for discussion). I also assume that for a form to be reconstructable to Proto-Tibeto-Burman, there should be a statistically significant representation of the form in the family (see Nichols 1996), and so I would not reconstruct a form to Proto-Tibeto-Burman based on the forms in two or four languages out of hundreds. In general, I would not leap to farfetched conclusions on the basis of

---

[2] According to the system of Fang-Kuei Li (1971), *lag would be the reconstruction for *yè* (夜) 'night'. The TB root is reconstructed as *s-gla in Benedict (1972:42), but corrected to *s/g-la in Matisoff (2003). Mei was using Li's reconstruction, though he also recognized the cognacy of *xī* 夕 *ljiak. These two forms are reconstructed by Baxter (1992) as *(l)jAks and *z(l)jAk, respectively.

one or two forms that do not even match semantically, such as Guillaume's conclusion (2010b, 2012) that, since the vowel in one verb root in Tibetan changes from -*a* to -*o* in the perfective, and in some Kiranti languages the vowel of some verb roots supposedly changes from -*a* to -*o* due to a third person -*u* suffix, and in Bantawa this vowel change only happens with past tense forms, then this is evidence that Tibetan has a person-marking system.

The solution to these epistemological problems is to be aware of the principle of interpretation and be aware of how assumptions can influence our interpretations, and try to work as inductively as possible, and make explicit arguments for all generalizations and assumptions.

3. Subgrouping and reconstruction in Sino-Tibetan

There are currently three problems with subgrouping and reconstruction work in Sino-Tibetan:

(1) Lack of good data, plus misuse of data that is there by people trying to make a particular point. An example of the latter is Guillaume's mention of languages in a way that makes it seem their data support his argument even when they do not, such as in citing Hayu in his conference paper in support of the *a > o* person marking argument, even though the facts of Hayu do not actually support the argument, or even talking about the supposed *-u suffix that supposedly causes this alternation as a 3sg patient marker in Kiranti, when Ebert (1990:58ff.) had shown conclusively that this is direct marking (in complementary distribution with the inverse marking, and used in 2>1 situations with no 3rd person referent involved), and not 3rd person marking. Another aspect of this is cherry-picking items from dictionaries to support some view while ignoring conflicting data in the same dictionary, often on the same page.

(2) Influence of shared geographic location on languages and on researchers' assumptions about relatedness. If we look at the various proposals for subgrouping in Tibeto-Burman, almost all include geographic designations such as "Western Himalayan". 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 into and/or through certain regions (see LaPolla 2001), 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. The debate about the use of the term "Kamarupan" hinges partly on the validity of using geographic terms and the related assumptions.

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 Bodish (Tibetan and its close relatives) and the rest of Tibeto-Burman was due to the early migration of those who became the Bodish speakers west and then south throughout Tibet and down into Nepal, Bhutan, and India (unlike the rest of the Tibeto-Burman languages, which either stayed in north-western China or migrated down the river valleys east of Tibet into Southeast Asia—see LaPolla 2001 for details). But some of the languages now classified as Bodic (a much larger grouping than Bodish), such as the Kiranti languages and what are called the Western Himalayan languages of Uttar Pradesh, show what I shall describe as "individual-identifying correspondences" with languages that clearly did not come down with the Bodish migration, such as Rawang. At the same time they share very little in the way of such correspondences with the Bodish languages.

The most likely explanation is that those so-called non-Bodish Bodic languages did not come down from the north, but came across from the (south-)east. They show similarities with the true Bodish languages because of contact rather than genetic inheritance. Non-linguistic evidence also points to this conclusion as well (see the papers from the conference "Origins and Migrations Among Tibeto-Burman Speakers of the Extended Eastern Himalaya" organized by Toni Huber and Stuart Blackburn, Humboldt Universität zu Berlin, May 23-25, 2008).

(3) Lack of consistent and clear standards and principles for subgrouping. That is, no consensus on methodology. Though there is some excellent work done using the comparative method, and there have been arguments for more rigorous application of the comparative method (using sets of unusual shared innovations—Thurgood 1982), subgrouping within Sino-Tibetan 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), or, as we saw above, because the languages just happen to be in the same geographic area.

4. An alternative methodology

The solution to these problems I think lies in finding a more scientific methodology. The paper Johanna Nichols did for the book *The Comparative Method Reviewed* (1996) is a discussion of what types of evidence we need to show relatedness between two languages (see also Lehmann 2005). 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" (Nichols 1996:48).

The reason for this is that there is very little likelihood of the entire paradigm appearing in different unrelated 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 proto-language. This type of evidence Nichols calls "individual-identifying" evidence, and she opposes it to "type-identifying" evidence, features such as a particular word order, or the simple presence of a particular category such as non-cognate gender or ergativity or conjunct-disjunct marking, 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 for the evidence to be taken as pointing to a single proto-language. This level is achieved by the evidence used for relatedness in Indo-European linguistics.

In monosyllabic morphology-poor languages, achieving this standard is more difficult, but not impossible. What is needed is for particular elements (words and/or morphological markers) to be organized into paradigm-like sets and applied rigorously to determine relatedness. That is, the *sets* are treated as if they were paradigms and used as individual-identifying evidence of relatedness, because the particular combination of independent elements as an internally-structured set would give us the level of statistical significance we need. 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. 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. Ikeda (2007) is an attempt at this using six words that are shared by all languages classified uncontroversially as Qiangic languages, and not by neighboring languages which are not uncontroversially Qiangic. He takes these six words as a set, and applies this set to the question of whether Tangut should or should not be considered a Qiangic language. That is a small set, but the results were interesting, and this method has also been used by Chirkova (2009) in discussing the affiliation of Shixing.[3]

---

[3] Burling (1982) also involves using a set of key lexical items to define what he calls the Sal grouping, but as the set of items was not applied rigorously as a set, the results are open to question (see Coupe 2012).

Without this rigorous use of the items as a set, simple 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 paradigm-like sets, or create sets of features. In a paper presented to the 33rd International Conference on Sino-Tibetan Languages and Linguistics in 2000, I created such a set using the person marking and reflexive/middle marking and certain other distinctive features found in a subset of Tibeto-Burman languages that can be taken together to identify the individual protolanguage from which those languages descended, and so identify that subset of languages within Tibeto-Burman as a valid genetic grouping. I later called this group "Rung". This is not a departure from the traditional comparative method; it is in fact simply applying the comparative method rigorously. Looked at in more traditional terms we are simply using shared innovations for identifying genetic groupings (the person-marking paradigm used is a clear innovation from the point of view of grammaticalization theory—LaPolla 1992), but we are being rigorous in saying that the same set of innovations must be manifested in all the languages in the same way.

Nichols' main goal in her paper is to set up a clear criterion for determining whether some feature is or is not useful in determining the relatedness of languages, and she 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 the following:

> 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. [Pp.60-61, emphasis in original.]

In this passage Nichols is talking about the level of attestation required of a piece of evidence for it to be considered evidence of genetic relatedness (i.e. be individual-identifying). That is, if we find a particular individual-identifying set of features in all the non-controversial branches of a recognized family, then that set can be used as a test for membership in the family. For Sino-Tibetan as a whole, we might be able to use the following set of features, if a survey of the languages showed each of these features did appear in all of the uncontroversial major ST languages as a set: the numerals, the voiced (or unaspirated) vs. voiceless (or aspirated) opposition in initial consonants marking intransitive vs. transitive or simplex vs. causative, the *s- causative prefix, the *-t transitivizing suffix, and the *-n and *-s formative suffixes. If established, the set could then be used to determine whether an unclassified language (or language group) is or is not Sino-Tibetan. We could also create such a set of isoglosses for Sinitic vs. Tibeto-Burman, using the fact that PTB reconstructs to a system with two series of stops, whereas Chinese reconstructs to a system with four series of stops, and could include the negator *ta, which only reconstructs to TB and not to Sinitic, and the *p-initial negators that Sinitic has but TB does not have. There are also lexical isoglosses that could be included in the set, such as *tin for 'sky' as opposed to PTB *r-məw.

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, we can easily decide whether the feature should or should not be reconstructed to the Proto-Tibeto-Burman (or Proto-Sino-Tibetan) level. For example, person marking in Tibeto-Burman only appears in about half of the lower level branches of Tibeto-Burman, and so does not reach the level of statistical significance required to be reconstructable to Proto-Tibeto-Burman. 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 Tibeto-Burman, as I did in my ICSTLL paper in 2000.

In Appendix 1, I have given the beginnings of a Master list of features that might be used for assessing the relatedness of ST languages. This is the initial step to see what features are widely found and which are restricted to certain language groups. Much more work needs to be done to complete the set.

In Appendix 2, I have given the beginnings of a chart for the Tibetan varieties, trying to determine whether the proposed set is actually attested in the branch in all or at least a statistically significant number of the languages. If so, then it could be evidence of their relatedness, and also evidence of the reconstructability of those features to the proto-language. Nicolas Tournadre (e.g. 2008) has been trying to do something similar for what he calls 'Tibetic'. He has been trying to come up with a set of features for determining if a language is Tibetic or not, and I had some discussion with him about this recently. He also argues that all the modern Tibetan languages or dialects share common phonological innovations as well as some basic lexical and grammatical items (not found in other TB languages), such as *rdo* 'stone' and *'dug* 'sit' in terms of lexical items, and palatalization of *ty, *ly, *sy, *tsy (already found in OT, e.g. *gcig* 'one') and *ml > *md* or *mj* (the latter after a front vowel) in terms of phonological innovations. He also argues that all the modern Tibetan languages or dialects derive from Old Tibetan, a language closely related to Classical Literary Tibetan.

In Appendix 3, I have given the beginnings of the Tamangic languages chart, based on Noonan 2007. It will be noticed that some of the items in this list are also included in the Tibetic list. The overlap of these features in Tibetic and Tamangic is what gives us the Bodish grouping. All of these sets, and others that will need to be created, will require much more work, but I think the outcome will be a lot more reliable than simply trolling through dictionaries looking for cognates.

Appendix 1: Master list of items for creating sets of features

	Sinitic	Tibetic	Tamangic	Kiranti	Rawang
*s- causative prefix	+	+		+	+
initial voicing alternation (trans/intrans)	+	+		+	+
*ma- negative	+	+	–	+	+
*-pa nominalizer	(–)	+	+	+	+
*-t transitivizing suffix	+	+		+	+
*ta- prohibitive	–	–	+		–
*p- negators	+	–	–	–	–
*-ŋa 1sg suffix on verbs	–	–	–	+	+
*-na 2sg suffix on verbs	–	–	–		–
*-si reflexive/middle marker on verbs	–	–	–	+	+
*-tsi dual suffix on pronouns and verbs	–	–	–	+	+
*ruŋ 'sit/exist'	–	–	–	+	+
marked event prefix	–	–	–	+	+
velar 1pl exclusive marker	–	–	–	+	–
-le LOC/DAT/PURP	–	–		(+)	+
-mi/i agentive	–	–	–	–	+
-səŋ LOC/DAT/PURP	–	–	–	–	+
-mkhan nominalizer	–	+	–	–	–
-sa nominalizer	–	+	–		–
-(ky)i genitive	–	+	–	–	–
-(kyi)s ablative/instrumental/agentive	–	+	–	–	–
-la allative/dative/locative/ (patient)/experiencer	–	+	+	+	–
dang comitative	–	+		–	–
*-s perfective	?	+			–
bdun 'seven'	–	+	–	–	–
rta 'horse'	–	+	–	–	–
(b)u diminutive	–	+	–	–	–
mi- negative	–	+	–	–	–
*yin copula	–	+	+	–	–
*ŋua copula	–	–	–	–	–
*mu copula	–	–	+	–	–
*ta 'become'	–		+		?
*-la conditional/irrealis	–		+		–
*ha- negative	–	–	+		–
*-ci perfective	–	–	+		–
*-si sequential converb	–	–	+		–
*-kay simultaneous converb/future	–	–	+		–
*-(k)u imperative	–	–	+		–
*-e interrogative	–	–	+		–

Appendix 2: List of items for establishing Tibetic group[4]

variety	*s-	-mkhan	*-pa	-sa	(ky)i	-s (erg)	-s(PFV)	bdun	rta	la	dang	yin
Classical	+	+	+	+	+	+	+	+	+	+	+	+
Lhasa	+	+	+	+	+	+	+	+	+	+	+	+
Mustang		+	+	+	+	+			+			+
Ngari		+	+	+	+	+			+			+
Sherpa	+	+	+	+	+	+		+	+	+	+	+
Shigatse		+	+	+	+	+			+			+
Lhomi		+	+	+	+	+			+	+		+
Jirel	+	+	+	+	+	+			+	+	+	+
Balti	+	+	+	+	+	+			+	+	+	+
Jad		+	+	+	+	+			+			+
Leh		+	+	+	+	+			+			+
Purki		+	+	+	+	+			+			+
Spiti		+	+	+	+	+			+			+
Tod		+	+	+	+	+			+			+
Ladakhi		+	+	+	+	+			+	+	+	+
Dzongka	+	+	+	+	+	+	+	+	+	+	+	+

---

[4] This chart was completed to the extent that it is with the help of Nicolas Tournadre. I wish to thank him for that. Boxes left blank are awaiting further information about the varieties.

Appendix 3: List of items for Tamangic

variety	*yin	*-pa	*mu	*ta	*-la COND	*ha-	*-ci PFV	*-si converb	*-kay	*-(k)u	*-e	*-sə ERG	*-r/la >GEN	dat/loc	'infin'	RECIP
Tamang		ba														
Gurung													-e	-r	-l(aʔ)	
Manange		-pʌ											-lʌ			
Nar-Phu		-pɛ	mʉ	tâ		a-						-se	-ye	re	-ne	
Takali																
Seke												-sə	-ye			
Chantyal	ɦin	-wa	mu	ta		a-		-si				-sə	-ye			-khum

128

References

Baxter, William H. 1992. *A Handbook of Old Chinese Phonology*. Berlin & New York: Mouton de Gruyter.

Benedict, Paul K. 1972. *Sino-Tibetan: A Conspectus*. Cambridge: Cambridge University Press.

Burling, Robbins. 1982. The Sal languages. *Linguistics of the Tibeto-Burman Area* 7.2: 1-32.

Chirkova, Katia. 2009. Shixing, a Sino-Tibetan language of South-West China: a grammatical sketch with two appended texts. *Linguistics of the Tibeto-Burman Area* 32.1:1-89.

Coupe, Alexander R. 2012. Overcounting numeral systems and their relevance to sub-grouping in the Tibeto-Burman languages of Nagaland. *Language and Linguistics* 13.1:193-220.

Driem, George van. 1993. The Proto-Tibeto-Burman verbal agreement system. *Bulletin of the School of Oriental and African Studies* 56.2:292-334.

Ebert, Karen H. 1990. On the evidence for the relationship Kiranti-Rung. *Linguistics of the Tibeto-Burman Area* 13.1:57-78.

Halliday, M. A. K. 1994. *An Introduction to Functional Grammar* (2nd edition). London: Edward Arnold.

Ikeda, Takumi. 2007. Qiangyuzhi yuyan de tezhengci: shi tan Xixiayu he Qiangyuzhi de guanxi [Characteristic words of the Qiangic languages: a contribution to the comparative study of Qiang and Tangut]. Paper presented at the 40th International Conference on Sino-Tibetan Languages and Linguistics, September 26-29, 2007. Harbin, China.

Jacques, Guillaume. 2010a. A possible trace of verbal agreement in Tibetan. *Himalayan Linguistics* 9.1:41-49.

Jacques, Guillaume. 2010b. Sino-Tibetan morphology from a Rgyalrong perspective. Paper presented at the International Symposium on Sino-Tibetan Comparative Studies in the 21st Century, June 24-25, 2010. Taipei: Academia Sinica.

Jacques, Guillaume. 2012. Agreement morphology: the case Rgyalrongic and Kiranti. *Language and Linguistics* 13.1:83-116.

LaPolla, Randy J. 1992. On the dating and nature of verb agreement in Tibeto-Burman. *Bulletin of the School of Oriental and African Studies* 55.2:298-315.

LaPolla, Randy J. 2000. Subgrouping in Tibeto-Burman: Can an individual-identifying standard be developed? How do we factor in the history of migrations and language contact? Paper presented at the 33rd International Conference on Sino-Tibetan Languages and Linguistics, October 2-6, 2000. Bangkok, Thailand. To appear in

What's where why? Language Typology and Historical Contingency, ed. by Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake. Amsterdam & Philadelphia: John Benjamins.

LaPolla, Randy J. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, ed. by Alexandra Y. Aikhenvald & R.M.W. Dixon, 225-254. Oxford & New York: Oxford University Press.

LaPolla, Randy J. 2003a. Overview of Sino-Tibetan morphosyntax. *The Sino-Tibetan Languages*, ed. by Graham Thurgood & Randy J. LaPolla, 22-42. London: Routledge.

LaPolla, Randy J. 2003b. Why languages differ: variation in the conventionalization of constraints on inference. *Language Variation: Papers on Variation and Change in the Sinosphere and in the Indosphere in Honour of James A. Matisoff*, ed. by David Bradley, Randy J. LaPolla, Boyd Michailovsky & Graham Thurgood, 113-144. Pacific Linguistics 555. Canberra: The Australian National University.

Lehmann, Christian. 2005. On the methodological bases of genetic language comparison. *Language Research* 41.2:379-404. Seoul: Seoul National University.

Li, Fang-Kuei. 1971. Shangguyin yanjiu [Studies on Archaic Chinese phonology]. *Tsing Hua Journal of Chinese Studies*, New Series 9.1-2:1-61.

Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction.* Berkeley: University of California Press.

Mei, Tsu-Lin. 1979. Sino-Tibetan 'year', 'month', 'foot' and 'vulva'. *Tsing Hua Journal of Chinese Studies*, New Series 12.1:117-133.

Nichols, Johanna. 1990. Linguistic diversity and the first settlement of the New World. *Language* 66.3:475-521.

Nichols, Johanna. 1996. The comparative method as heuristic. *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, ed. by Mark Durie & Malcolm D. Ross, 39-71. New York: Oxford University Press.

Noonan, Michael. 2007. Nominalizers in Tamangic languages. Paper presented at the International Workshop on Nominalizers and Copulas in East Asian and Neighboring Languages, January 8-12, 2007. Hong Kong: The Chinese University of Hong Kong.

Thurgood, Graham. 1982. Subgrouping on the basis of shared phonological innovations: a Lolo-Burmese case study. *Berkeley Linguistics Society* (*BLS*) 8:251-260. Berkeley: Berkeley Linguistics Society.

Tournadre, Nicholas. 2008. The notion of scale in linguistic classification: Is 'Tibetan' a language or a family of languages? Paper presented at the 14th Himalayan Languages Symposium, August 21-23, 2008. University of Gothenburg, Sweden.

Watkins, Calvert. 1976. Towards Proto-Indo-European syntax: problems and pseudo-problems. *Chicago Linguistic Society* (*CLS*) 12:305-326. Chicago: Chicago Linguistic Society.

Linguistics
La Trobe University
VIC 3086 Australia
r.lapolla@latrobe.edu.au

論漢藏語比較語言學的方法和依據

羅仁地

拉籌伯大學

　　本文探討有關漢藏語系語言之間親屬關係研究以及這些語言特徵構擬的認識論和方法論問題。其中主要的問題之一是我們沒有公認的標準，尤其是能經得起統計概率的考核。本文提出一種新的方法，即使用一套固定的特徵來衡量一些語言的親屬關係。這種方法能夠讓我們計算這些語言有共同來源的統計概率。

關鍵詞：漢藏語，歷史比較法，統計分析，原始語言構擬方法