Once again on methodology and argumentation in linguistics

Problems with the arguments for recasting Sino-Tibetan as “Trans-Himalayan”

Randy J. LaPolla
Nanyang Technological University, Singapore

There have been challenges to the received view of the structure of the Sino-Tibetan language family. This is all well and good, as we should constantly challenge our most basic assumptions. In this paper I look at the arguments presented with a view to convincing us to change our conception of Sino-Tibetan and to change the name of the family to “Trans-Himalayan”, and find them less than convincing, due to problems of fact and argumentation.

Keywords: Sino-Tibetan linguistics, historical linguistics, argumentation, Trans-Himalayan, Chinese linguistics, Tibeto-Burman linguistics

1. Introduction

In a series of papers over the last 26 years I have been arguing for more empirically based, more typologically and theoretically informed, and more logically rigorous analyses of phenomena in Sino-Tibetan historical linguistics and linguistic typology. In LaPolla 1990, 1993 (see also LaPolla & Poa 2006 and my papers on information structure in Chinese), by taking a more empirical approach towards the

---

1. Personal note: I moved from philosophy into linguistics because although the latter had rigorous argumentation, there was no empirical “bottom line” that one could rely on in deciding between differing approaches. Unfortunately, at that time (1980’s) linguistics was in the grip of Chomskyan rationalist philosophy, and many non-Chomskyan works were riddled with problems of argumentation as well. The papers mentioned here have been a small attempt to move linguistics back towards a more empirical and rigorous direction.
analysis of Chinese\textsuperscript{2} I presented evidence that questioned the universality of the grammatical relations “subject” and “direct object” and showed that the typology of alignment systems needs to include languages with no grammatical alignment, such as Chinese, which has a purely topic-comment structure.\textsuperscript{3}

In LaPolla 1992a I raised several serious doubts about the theoretical and methodological validity of reconstructing person marking on the verb to Proto-Tibeto-Burman, and at the same time tried to show how using functionally and typologically based theories of grammar could benefit us in our work. I argued that in doing reconstruction work, we need to first do internal reconstruction to peel back the layers of grammaticalization, and not reconstruct transparent grammaticalizations to the proto-language, and also showed how looking at only a small amount of skewed data can lead to misanalyses, such as analyzing the hierarchical person marking systems in some Tibeto-Burman languages as ergative, as had been done because of only using examples with third person actors (e.g. Bauman 1979, see also DeLancey 1989 on the system as ergative). I also pointed out the difference between zero marking (which necessarily involves paradigmatic oppositions) and simple lack of marking, which is the pattern more commonly found in Tibeto-Burman, as the relational marking in many languages is not paradigmatic, but used when the speaker feels it is needed for clarity. In this regard I (1992b) also showed how Matthew Dryer’s (1986) assumption that what he called “Primary Object” marking was a grammatical relation in all languages that manifested the relevant pattern (where the patient and the recipient can be marked the same) did not hold up when the actual use of the relevant forms in Tibeto-Burman languages was considered. Instead I argued it is semantically motivated marking used when there was possible confusion between an agent and a non-agent, so I called the marking “anti-ergative” and later (1994a) “anti-agentive” marking.

In LaPolla 1995a I showed that, contrary to what had been assumed up to that time, ergative relational marking was not reconstructable to Proto-Tibetan-Burman or even to middle level proto-languages other than Proto-Bodish, and also argued there again for the difference between systemic and non-systemic relational marking, which has now become the hot topics known as “differential A(gent) marking” and “differential O(object) marking”. In LaPolla 1995b I showed

\textsuperscript{2} Solid empirical and insightful work on Chinese had been carried out by Y. R. Chao (e.g. 1968), but this work was either misunderstood or ignored, and most work on the language has relied on made-up sentences forced into preconceived notions of how a language should work rather than on careful inductive analysis.

\textsuperscript{3} Later work, e.g. LaPolla & Poa 2006, showed that Tagalog and other languages of the Philippine type also need to be recognized as a separate alignment type, and not forced into an ergative or accusative straightjacket.
how the theoretical concepts of markedness and prototypes can be used to help us understand how the morphological systems developed over time. In that paper, and in LaPolla & Yang 1996 and LaPolla 1996, inspired by insights gained from doing close, typologically informed fieldwork and analysis of Dulong data, I pointed out the existence and development of middle voice marking in Dulong, and later showed it to be found in other languages in the Tibeto-Burman family as well.

In 1994b I argued that we must be more rigorous in our search for cognates in historical work; while we can recognize variation, the recognition of variation in an initial or final must be based on clear correspondences in the other elements of the form. In LaPolla 2000 (published as 2013a; see also 2003) I showed how the common practice of using random lists of words to establish genetic relations within the family leads to statistically problematic results, and argued for the methodology followed by the Neogrammarians (Nichols 1996, see also Campbell 2003), using paradigms of morphology as evidence in determining genetic relatedness, as they allow for a more statistically significant and less arbitrary outcome (see also Ringe 1992, 1995, 1996, 1999). In LaPolla 2001 I brought in the demographic perspective that in order to properly understand the history of the languages in Sino-Tibetan one must first understand the history of the speakers of the languages, particularly the migrations, and showed how the migrations that have been recorded help us to understand why the family is the way it is. Ignoring this history leads to problematic hypotheses about genetic relatedness and the homeland of the family.

In LaPolla 2002 (originally written for a conference in 1994) I brought up problems with the methodology, argumentation, and explanations common in word order typology studies, and argued for a more logically and empirically rigorous approach that respects the facts of the languages involved, and this was followed up by LaPolla & Poa 2006, which took the position that the designations “SVO”, “SOV”, “OSV”, etc. were problematic non-empirical generalizations, and so in describing word order we should instead describe the actual principles that determine the word order in the language, and used the differences between Chinese, English, and Tagalog in terms of these principles as examples.

In LaPolla 2006a I argued that as linguistics is the study of linguistic structure and its use, and as structure emerges from use, in order to understand why languages are the way they are, we need to look at language in actual use, taking into account the entire communicative situation, and also appreciate the diversity of language structures, and not force languages into preconceived categories, and followed this with a paper in 2007 on how to do fieldwork in a way that allows the language’s categories to be discovered rather than having preconceived categories imposed on the language. In LaPolla 2008a (published in 2014) I applied this inductive methodology to the analysis of a Tagalog text and found that Tagalog
has clear grammaticalized constructions, but they do not correspond with noun phrase and verb phrase in other languages, so imposing categories such as noun phrase and verb phrase on the language, as had previously been done, is problematic.

Early on I had argued for a more empirically based constructionist approach to grammatical relations (e.g. Ch. 6 of Van Valin & LaPolla 1997, also LaPolla 2006c-d), but in 2008b (published as LaPolla, Kratochvíl & Coupe 2011) extended this insight to the study of transitivity, and later (2013b) to form classes as well.

In 2012a, following up on the 2000 paper on methodology in determining genetic relations, I pointed out the epistemological problem that scholars often take the language they know best as the most archaic in the family, and reconstruct the proto-language to look like that language (“Teeter’s Law”); the problem that scholars often cherry-pick their data, ignoring counter-evidence; and the problem of shared geographic location influencing studies of genetic relatedness, as languages in the same area are often simply assumed to be closely related. That latter view is problematic because it ignores the history of the waves of migration into most areas of the family and the contact-induced similarities that resulted. A third problem I pointed out there is the lack of a rigorous methodology that allows us to show the statistical probability of the identified cognate elements being unique to those languages. I then presented again the methodology first outlined in 2000, using morphological paradigms rather than random word lists for determining genetic relationships.

In 2012b I returned to the question of person marking in Tibeto-Burman, pointing out that the papers published against the view I presented in 1992a and elsewhere had not dealt with the main issue, which is that we should not reconstruct transparent grammaticalizations to the proto-language, and also marshaled other evidence to show that the relevant person marking system (there are many independently developed systems, see also LaPolla 1994a, 2001) on factual and theoretical grounds should not be reconstructed to the PTB level.

In 2016 I returned to the question of establishing a more empirical linguistics, arguing against non-empirical structuralist approaches to linguistics (2016a) and against the assumption in typology that one can classify a language in a particular category for comparative purposes even if the language does not manifest the relevant characteristics of that category (2016b); and arguing for a more data-focused empirical approach to Sino-Tibetan linguistics (2016c). In the present paper I return to the issue of logically rigorous argumentation, using arguments put forward for restructuring the Sino-Tibetan family and changing the name of the family to “Trans-Himalayan” as a case study.
2. A case study

In a series of articles and talks over a period of a dozen years (e.g. van Driem 2002, 2014), George van Driem has been arguing that the generally accepted view of Sino-Tibetan (with the highest level branching being between Sinitic and Tibeto-Burman, e.g. Benedict 1976, Bradley 1997, 2002, Matisoff 2003, Thurgood, in press) is not correct, as he says the Sinitic languages do not form a branch in opposition to the Tibeto-Burman languages, and argues we shouldn’t use the name “Sino-Tibetan”, but should use the name “Trans-Himalayan” for the family, as he argues the homeland of the family is the Himalayas. It is certainly good practice to challenge even our most deeply held assumptions, especially if new data can be produced or clear analytical arguments based on facts can be made.

In 1997 van Driem produced data to try to support his view at that time of “Sino-Bodic”, arguing for a closer relation between what he called Bodic (essentially Limbu and other Kiranti languages) and Sinitic, but this paper was fully refuted by Matisoff (2000) on methodological and factual grounds. Van Driem has not accepted this refutation (2005, 2014: 23), but since then has not produced any data to support his claim that might be scientifically evaluated. He has instead relied on problematic argumentation. I will go through the arguments in van Driem’s most recent article on this (2014), with the goal of stimulating serious reflection on what constitutes a valid argument in historical linguistics.

1. Van Driem’s main argument is that in 1823 Julius Klaproth, a self-taught Asianist, identified a language family, which van Driem says was later named “Tibeto-Burman”, that saw Chinese, Tibetan, and Burmese as sister branches on the family tree. Van Driem states that Klaproth had got it right, and so we should go along with Klaproth’s view. Van Driem does not give any evidence for Klaproth’s Tibeto-Burman family or even any clear indication of what it included other than “Burmese, Tibetan and Chinese and all of the languages which could be demonstrated to be related to these three” (2014: 12), though van Driem says, “He explicitly excluded languages today known to be Kradai or Daic (e.g. Thai, Lao, Shan), Austroasiatic (e.g. Mon, Vietnamese, Nicobarese, Khmer) and Altaic (e.g. Japanese, Korean, Mongolic, Turkic)” (ibid.).

---

4. See also Benedict 1976 for refutation of the suggestion of a supposed closer relationship between Sinitic and Tibetan. My own view of such suggestions is that they are due to a bibliographic bias: as Chinese and Tibetan have the most extensive and oldest dictionaries, those who rely on random searches in dictionaries for cognate words in determining genetic relationships will find it easier to find supposed cognates in dictionaries of these two languages than any other.

5. Only this work by Klaproth (i.e. 1823) is cited by van Driem in the 2002 and 2014 papers.
There are three problems with this:

The first is that I cannot find any statement in the book by Klaproth that van Driem cites arguing for such a grouping, as van Driem claims. Klaproth simply provides lexical lists and mentions similarities or the lack thereof, without any systematic comparisons.

Second, assuming Klaproth did make such a claim elsewhere, no supporting evidence is given by van Driem as to why we should go along with Klaproth’s view. Klaproth was working in the early decades of the 19th century, before the creation of the Neogrammarian comparative method (which itself was not used to determine genetic relationships — see Nichols 1996). Klaproth had no methodology other than looking at words that appeared similar in the various languages, and he has a list showing commonalities among all languages (Klaproth 1823: 36–39). Because he was just working with similarities in modern forms, Klaproth saw words like Chinese bí 鼻 (transcribed incorrectly with a voiced initial) ‘nose’ and Persian bini ‘nose’; German Ohr (transcribed uhl, uhr) ‘ear’ and Chinese ěr 耳; and Chinese pán (transcribed as p’an) and German Pfanne ‘pan’ as related (he saw general commonalities and more specific relationships). If a modern scholar uses this methodology we laugh him out of the room, but van Driem is asking us to take this seriously.

Third, as Klaproth had no methodology other than matching modern forms, how can the expression “all of the languages which could be demonstrated to be related to these three” used by van Driem be evaluated? In an otherwise very positive biography of Klaproth, Walravens (2006: 181) commented, “Klaproth’s talents were not so much creative but critical. He had an amazing command of languages but his main linguistic contributions were lexicographical, not comparative or grammatical.” The German Wikipedia article about von Klaproth (http://de.wikipedia.org/wiki/Heinrich_Julius_Klaproth) states: “Sein wissenschaftliches Werk galt bereits 1911 (Encyclopaedia Britannica) als völlig überholt und ist nur noch von literarischem Interesse.” [His scientific work counted already in 1911 (Encyclopaedia Britannica) as completely outdated and is only of literary interest.] If this is so why are we to take him as an authority? And if we do, do we also take seriously his discussion of the languages before and after the Biblical flood (1823: 42)?

The only larger grouping that Klaproth actually mentions is what he calls the “Trans-Gangetic” languages (Transganetischen Sprachen). The list of these languages given by Klaproth (pp. 367–405) includes Vietnamese, Thai, and Burmese, so van Driem’s statement that Klaproth, “… explicitly excluded languages today known to be Kradai or Daic (e.g. Thai, Lao, Shan), Austroasiatic (e.g. Mon, Vietnamese, Nicobarese, Khmer) …” (2014: 12) is also problematic.
We can acknowledge that for his time Klaproth was quite an interesting person and may have made some contributions to the development of the field at that time, but to assume that he had access to the same quality of data and methodological tools that we have access to now and so could make a better assessment of the proper subgrouping of the Sino-Tibetan languages than we can now is very problematic.

For all these reasons we should not take Julius Klaproth as an authority on Sino-Tibetan subgrouping, and if that is van Driem’s only evidence for the proposed view, then we cannot take that view seriously. Aside from the problems with Klaproth’s status as an authority, there is also the problem that in logic this sort of appeal to authority rather than to evidence would be considered a fallacious type of argumentation, argument from authority (argumentum ab auctoritate).

2. A second theme of van Driem’s articles (e.g. van Driem 2002, 2014) seems to me to be an ad hominem attack (another type of fallacious argumentation) on those who support the current view of Sino-Tibetan. Van Driem goes through a history of the development of ideas about language subgroups, and tries hard to associate certain views, such as our current understanding of Sino-Tibetan, with the racist views of certain German typologists in the 19th century. In doing this it seems to me he is insinuating that those who subscribe to the standard view of Sino-Tibetan are racist. If that is his intention, it is of course ridiculous.

First, the racist views of the typologists actually had little to do with genetic subgrouping, but were related to the typological classification of languages. Also, since we are talking about the racist views of early 19th century scholars, why leave out Klaproth’s view (1823: 344) that the Tibetans look like monkeys?

Second, those of us who subscribe to the standard view of Sino-Tibetan do so not because of influence from some early 19th century scholars, but because of Prof. Li Fang-kuei’s analysis of the languages of China in his 1936–37 article and also later comparative work by Benedict (e.g. 1972, 1976) and Matisoff (e.g. Matisoff 1973, 2003) modifying Li’s original view. Van Driem is either unaware of Prof. Li’s article and its influence or simply choses not to consider any Chinese scholars in talking about this issue so that he can make his connection with racism.

3. A third theme of van Driem’s articles is his view that the case hasn’t been made for the standard view of Sino-Tibetan, and so we should accept Klaproth’s view. This ignores all the work done within the standard view, but even if it were true that the case for Sino-Tibetan hasn’t been fully proven, this is an argument from ignorance (argumentum ad ignorantiam), and shifting the burden of proof (onus probandi), both fallacious argumentation types. Van Driem criticizes Benedict for “isolating Chinese as the odd-man out” (2014: 15), and not seeing Sinitic as part of Tibeto-Burman, and says “… no evidence has ever been adduced in support of the
Sino-Tibetan phylogenetic model, defined by its truncated ‘Tibeto-Burman’ taxon encompassing all non-Sinitic languages … All comparative evidence amassed to date supports Julius von Klaproth’s 1823 minimalist Tibeto-Burman tree, which epistemologically therefore continues to represent the default model.” (p. 16). This is problematic in a number of respects.

It ignores the fact that the debates in the early days of the Sino-Tibetan conferences and discussion of Benedict’s Conspectus (1972) were often about whether Chinese belonged with the Tibeto-Burman languages at all (see the discussion in Benedict 1976), showing how different Chinese was thought to be, and the fact that Benedict (1976) showed not only that Sinitic should be part of Sino-Tibetan, but that the first bifurcation of the family should be between Sinitic and Tibeto-Burman.6 See also the discussion of isoglosses distinguishing TB from Sinitic in LaPolla 2012a. The view that the initial split was between Sinitic and the rest of the languages is also supported by the history of migrations. As discussed in LaPolla 2001 and references therein, there was an initial split between the Sinitic and non-Sinitic varieties due to the different directions of their migrations, and then the Bodish languages broke off from the rest of Tibeto-Burman, as they migrated west and south into Tibet, while the rest of the Tibeto-Burman languages (in different waves) followed the southern route down the river valleys and around the southern edge of the Tibetan plateau.

In terms of the logic of the argument, if it really were the case that the bifurcation hadn’t been argued for, the fallacy would be in saying that there is no evidence and so it must not be true. That is a problem in itself, and van Driem’s statement in support of what he says is Klaproth’s view is also problematic as there is no evidence for what he says is Klaproth’s tree, so how can he say that not having evidence of a particular view is proof that it is wrong? This is self-contradictory.

4. Van Driem says (2014: 16) that “Trans-Himalayan” is a better term than “Sino-Tibetan” because geographic terms are better than genetic designations. I (LaPolla 2012) have argued the opposite, that geographic terms ignore the history of migrations, implying that all the languages in that area have a particularly close relationship, and so prevent us from understanding the history of the languages.

Van Driem proposes the term “Trans-Himalayan” because he claims the family “straddles the great Himalayan range along both its northern and southern flanks” (2014: 16). He claims that “[b]y far most of the roughly 300 different Tibeto-Burman languages and three fourths of the major Trans-Himalayan subgroups are situated along the southern flanks of the Himalayas (Figure 3)” (2014: 16).

6. As late as 1990 Sagart was not convinced Sinitic formed a family with Tibeto-Burman, as he felt it was closer to Austronesian, though he now feels “ST as a whole, not just Chinese, forms a genetic unit with Austronesian” (2006: 208).
This is a very misleading image. It treats large families, like Sinitic, and what are essentially single languages, such as Nungish, as equal in this view, and this makes it seem like there is less diversity in areas covered by the larger yet internally very diverse groups than in the areas where there are many small groups. David Bradley long ago pointed out the sociological fact that within the Indosphere there is maximal differentiation of language varieties into languages, such that closely related mutually intelligible varieties are considered different languages, while in the Sinosphere there is minimal differentiation, and so many mutually unintelligible varieties are lumped together as languages (Bradley & Bradley 2002, Bradley 2015; see also Poa & LaPolla 2007 on the effects of the latter tendency on language maintenance in China). An example of the sort of diversity we can find is Pelkey 2008, 2011, which document the existence of 24 mutually unintelligible languages within a small area on the China-Vietnam border, all originally considered part of a single dialect of the Yi language. There is in fact much greater diversity north-east of the Himalayas, as we would expect, given the origin of the family there. None of the carefully worked out proposals for subgrouping in Tibeto-Burman (e.g. Benedict 1972, Bradley 1997, 2002, Matisoff 2003, Thurgood, in press) have such a plethora of micro-subgroups on the southern side of the Himalayas. See also Chappell 2015 (among others) on the diversity of Sinitic languages, showing how unrepresentative having a single dot for an entire branch of the family is.

Van Driem’s view also ignores the archeological and genetic record showing that relevant human habitation in the Tibetan plateau and the Himalayas
was relatively late and came from the Yellow River valley (Aldenderfer 2007, Aldenderfer & Zhang 2004, Brantingham et al. 2007, Chen et al. 2015, Rhode et al. 2007, Su et al. 2000). The move south of the Himalayas was relatively late in the overall spread of the family (LaPolla 2001 and references therein), so to call the family “Trans-Himalayan” is misleading both historically and in terms of current distribution. I therefore see no valid argument for changing the name of the family.

5. Van Driem proposes what he calls a “fallen leaves” model of the family, many language groups which supposedly happened to fall where they are, with no necessary connections among them (van Driem 2014: 19, Figure 5):

![Figure 2. Figure 5 of van Driem 2014, p. 19](image)

This is not a subgrouping model at all. It also ignores what we know of the migrations of the peoples. That is, they didn’t just “fall” where they are; there were clear migrations, often several waves into the same place, and this has affected the distribution and form of the languages. It also ignores all of the careful comparative work that has been done to show higher level relationships among some of these groups.

As David Bradley pointed out to me (email Jan 12, 2015), the fallen leaves model “implies the (former) existence of a tree but does not attempt to find it, which is not comparative/historical linguistics. If one breaks subgroups down very finely in some areas, then the density of subgroups will appear to be higher there; or conversely the ecology of some areas may be more conducive to formation and expansion of large creoloid groups (Bodo-Garo as Burling has proposed, DeLancey extends this to Burmic as well) and so the diversity may be lower than elsewhere — the prime example is of course Sinitic, which has expanded across
most of China from the upper Yellow River in the last four millennia and lost most of its hard edges in the process.”

6. Van Driem (2014) then goes on to talk about genetic evidence of population dispersions. This is irrelevant to the question of linguistic subgrouping, as archeological or genetic evidence does not tell us anything about the languages spoken at that time and is of a time depth much greater than what we are talking about in terms of the events that led to the current language distribution. Van Driem acknowledges these two points, but still goes on to make speculative connections between genes and languages.

What can be done with genetic evidence is of a negative nature, for example, if we could show that the Burmese speakers in southern Myanmar were genetically closer to the Mon people than to the northern Burmese, then it would be supporting evidence for the idea that there was no massive migration of Burmese speakers from the north into the south after the northern conquest of the south in the mid 18th century, but that the original Mon speakers simply switched to speaking Burmese (Bradley 1980).

3. Conclusion

In the history of science it has been the case that generally only overwhelming evidence of anomalies presented by a paradigm and a viable alternative can cause a change from one scientific paradigm to another (Kuhn 1970). In this case I do not find any evidence or a better alternative that would require me to rethink my understanding of the Sino-Tibetan family (LaPolla 2006b) or choose another name for the family.

Acknowledgements

This paper was presented at the 47th International Conference on Sino-Tibetan Languages and Linguistics, held in Kunming in October, 2014. I would like to thank all those who commented on the paper at that time, and thanked me for writing it and urged me to publish it, and also David Bradley, Alec Coupe, and Graham Thurgood and the anonymous reviewers for comments since then. I would also like to thank Yvonne Tse Crepaldi, Stefanie Stadler, and Rik De Busser for confirmation of my reading of the German text.
References


LaPolla, Randy J. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In R. M. W. Dixon & A. Y. Aikhenvald (eds.), *Areal diffusion*


LaPolla, Randy J. 2008a. Constituent structure in a Tagalog text. Keynote presentation to the 10th Philippine Linguistics Congress, University of the Philippines – Diliman, Quezon City, December 10-12, 2008. (Published as LaPolla 2014.)

LaPolla, Randy J. 2008b. Questions on transitivity. Keynote presentation to open the Workshop on Transitivity, Research Centre for Linguistic Typology, La Trobe University, 21 August. (Revised version published as LaPolla, Kratockvíl & Coupe 2011.)


LaPolla, Randy J. 2016b. On categorization: Stick to the facts of the languages. Invited position paper for special issue of *Linguistic Typology* 20 (2) on descriptive vs. comparative categories, in press.


LaPolla, Randy J. & Yang Jiangling. 1996. Dulong/Riwangyu dongci de fanshen he zhongjianzai biaozhi (Reflexive and middle marking in Dulong/Rawang). In Dai Qingxia et al (eds.), *Zhongguo minzu yuyan luncong (1)* (Collected essays on Chinese minority languages, 1), 13–34. Central University of Nationalities Press. (Published in English as LaPolla & Yang 2005.)


**Author’s address**

Prof. Randy J. LaPolla  
Division of Linguistics and Multilingual Studies  
HSS 03–45  
Nanyang Technological University  
14 Nanyang Drive  
SINGAPORE, 637332  
Singapore  
randylapolla@ntu.edu.sg