

On Principles and Practices of Language Classification

Katia Chirkova
CRLAO, CNRS

This article is an overview of issues in language classification, in particular in connection with three subgroups of the Tibeto-Burman language family: Tibetic, Sinitic and Qiangic. First, I discuss the practical application of currently prevalent classifications of Tibetic, Sinitic and Qiangic languages. Then, with reference to insights from classification practices in biology, I review alternatives for a practical classification in linguistics.

Key words: language classification, genetic classification, Sinitic, Tibetic, Qiangic, Tibeto-Burman

1. Introduction

It is a great privilege and pleasure to contribute to the Festschrift in honor of Alain Peyraube. I gladly and admiringly join this celebration of his scientific achievements, the diversity of his interests, his renowned erudition, and his foresight on issues of great impact to the field. In this article, I reflect on one of Alain Peyraube's most recent interests: linguistic classification from a general and more specific, East Asian, perspective. His contributions to the issue of classification (Peyraube 2005, 2007) touch upon major trends in the field of East Asian linguistics, namely:

- (1) unabating interest in classification issues, from periodic revisions of the precise composition of traditional groupings of languages in China and East Asia (such as Sino-Tibetan, Austronesian or Austro-Asiatic) to the construction of new macro-families (such as Austric, Austro-Tai, Sino-Tibetan-Austronesian or Proto-East-Asian)
- (2) strong emphasis on interdisciplinary studies, whereby linguistics is coupled with other disciplines (such as genetics, evolutionary biology, archaeology or anthropology) to test relevant phylogenetic hypotheses
- (3) fruitful exchange between biology and linguistics, inspired by a deeply entrenched view of parallelism between the development of living organisms and the development of languages

These major trends guide and shape daily linguistic practice. Since classification is fundamental to all aspects of linguistic research, my main concern in this article is the practical application of existing classifications and the impact of the aforementioned trends on ongoing synchronic investigations. Given my area of expertise, the ensuing discussion will be limited to the field of Tibeto-Burman languages.

2. The theory of evolution and genetic classification in linguistics

In a metaphor that is nearly as old as linguistics itself, language is often viewed as an organism. Ranging from a direct equation¹ to a more conventional symbolic portrayal, the conceptualisation of language as an organism has long accompanied the development of the discipline.² Indeed, parallels between languages and organisms abound (e.g. Janda & Joseph 2003:59-81, Peyraube 2007), allowing for profitable exchange between linguistics and biology.

Fundamentally, both biology and linguistics are grounded in the theory of evolution, which both disciplines hold to be the most adequate and unitary hypothesis to which a great variety of both biological and linguistic phenomena can be related. As a result, the most widespread and influential type of classification in linguistics is genetic, based on the assumption of common ancestry of languages and using basic vocabulary, sound correspondences and, whenever available, grammatical (essentially morphological) evidence as classification criteria. This type of classification is so basic to linguistics that no explicit justification of its validity and legitimacy is normally deemed necessary. Characteristically, the recent edition of the *Encyclopaedia of Languages and Linguistics* (2006) names genetic classification “the most satisfying way to group languages” for “professional linguists and general readers” (Blake 2006:446), but does not explain what is so satisfying about it. One of the few scholars to explicitly address this issue, Michael Noonan (2008:4), points out that genetic classification is not always useful in providing information about the structure of a language, especially on higher taxonomic levels. Instead, he notes, information about where in the world a language is spoken provides more useful information about grammatical structure. On the other hand, Noonan argues, genetic classification has proven a boon to historical linguistics and theories of language change. Ultimately, Noonan concludes, the most satisfying aspect of genetic classification is its sheer familiarity, since “What family does it belong to?” is typically “the first thing a linguist will ask on being told of an unfamiliar language.”

¹ Recently, language as a parasitic species (Kortlandt 1985, 2003, Mufwene 2001) or language as a symbiont (van Driem 2004, 2008).

² See van Driem (2008) for an overview.

In other words, genetic classification is popular due to its familiarity, but its practical value with regard to concrete linguistic data is questionable. Nonetheless, genetic classification is considered superior to other types of linguistic classification, for example, those based on typological features (see §3 for examples). This assumed superiority arguably results from the fundamental assumption that a classification based on evolution, which is in turn responsible for the existence and structure of natural languages, is imbued with explanatory power. Due to the deeply entrenched connection between explanation and theory in science, explanatory (viz. theoretical) analyses and frameworks are routinely valued higher than descriptive ones (cf. Dryer 2006:212-214).³

The dominance of genetic classification is the result of the sweeping success of Indo-European historical linguistics, which, in turn, is due to the clear evidence of a common origin for groups of languages in Europe (e.g. Romance languages) and the extensive amount of written records. The main metaphor that is used by Indo-European linguistics to explain the historical relationships, the genealogical family tree, is based on the assumption that the phylogenetic relationship between groups of languages is analogous to the genealogical relationship between individual languages, i.e. descent along a single genetic line (inheritance or vertical transmission).⁴ The straightforwardness of the model and its appeal in the Indo-European context stimulated its export to other linguistic areas.

Uncritical application of the family tree model to languages outside the Indo-European family has repeatedly engendered criticism, most recently and forcefully in Dixon (1997) or Aikhenvald & Dixon (2001). The conclusion of these two studies is that the family tree model is appropriate and can be proved through reconstruction in many cases (as in the case of Indo-European languages), but is problematic or even unsuitable in those cases where historical evolution is obscured by contact-related phenomena (as in the case of Australian languages). In other words, linguistic convergence and areal diffusion obscure historical development, so that originally genealogically unrelated languages may develop surface resemblances, forming groups which are synchronically similar, but genetically polyphyletic (that is, with multiple ancestors).

Borrowing and contact-induced change (i.e. horizontal transmission) as factors affecting the linear evolution of languages (i.e. vertical transmission) are now commonly taken into account in mainstream historical linguistics. Nonetheless, the impact of contact on linear descent remains subject of heated debate in linguistics, as in biology (cf. Dagan & Martin 2006:118.1). At present, awareness of the importance to enrich the

³ In the words of Karl Popper (1969:103): “The scientist aims at finding a true theory or description of the world [...], which shall also be an explanation of the observable facts.”

⁴ The drawbacks of the group to an individual analogy have been pointed out, among others, by Gilmour (1940:469-470) for systematics and, recently, Mufwene (2001:16-21) for linguistics.

dominant family tree model by taking into account “all the processes that affect language formation and development” (Chappell 2001:354) is on the increase. Much research is being undertaken to find objective criteria for assessing the balance between the effects of vertical and horizontal transmission on language change. Some of the most acute challenges include distinguishing loans from cognates, and common innovations from retentions.⁵ Attempts to “undo” the effects of contact include a methodology for separating innovations resulting from genetic inheritance from those diffused through contact, and a fully articulated theory of the relative diffusibility of features in a contact situation (e.g. Curnow 2001, Dench 2001). To date, all these attempts have been proven futile (Curnow 2001, Dench 2001) and are even argued to be doomed to failure (Thomason 2000).

In those cases where the linear development is known to be obscured by contact-related phenomena, as in the case of many East Asian languages (e.g. Benedict 1972, Matisoff 2001), and where the horizontal dimension of language change is supposedly as important for a coherent assessment of data as the vertical dimension, the emphasis is invariably on the latter (*viz.* genetic descent). The large number of relevant publications characteristically reflects the dominance of the genetic approach and the associated genetic classification. Also symptomatic of this dominance is the fact that general studies in language classification are limited to genetic classifications (e.g. Ruhlen 1994, Campbell & Poser 2008). Such studies mostly focus on higher taxonomic levels: families and macro-families. While major debates revolve around distant genetic relationships of considerable time depth (such as Nostratic or Amerind), classification issues at more empirically bound lower taxonomic levels have all along attracted less attention. This is remarkable, because it is at these lower taxonomic levels, where data are most voluminous and versatile, that the need for meaningful classifications to organize and store empirical knowledge is at its greatest. I have been confronted with the practical application of these low level classifications of languages (Tibetan and Chinese dialects (or Tibetic and Sinitic languages) and Qiangic languages) and will focus on this issue in the main part of this article.

⁵ A relevant example from the field of Tibeto-Burman linguistics is the Bai language, whose genetic affiliation has long been the subject of scholarly speculation. The ongoing controversy is due to the so far unresolved issue of whether numerous lexical similarities between Chinese and Bai reflect long-term contact between the two languages (loans) or inherited features (cognates).

3. Classifications of Tibetan and Chinese dialects and of Qiangic languages

Classification is a highly practical activity. It provides a practical means by which objects can be identified and compared, and the knowledge about them organized. In addition, a meaningful classification enables the researcher to make inductive generalizations concerning the classified data and to discover new knowledge of the variation and distribution of the examined attributes. In the words of Ernst Mayr (1976:427), a good classification, like a good scientific theory, “has a high predictive power with respect to the assignment of newly discovered species and the pattern of variation of previously unused characters. That classification is the best which is at least affected by such new discoveries”.

How do currently prevalent classifications of Tibetic, Sinitic, and Qiangic languages square with this desideratum for classification? I deal with this question as a practitioner who would like to use existing classifications as a descriptive tool and as a basis for inductive generalizations, while profiting from their predictive power with respect to adding new dialects or languages to the existing classification schemes.

Importantly for the ensuing discussion, the three groups considered (Tibetic, Sinitic, Qiangic) can be divided into two groups, based on whether the genetic relationship among the languages within each group is presupposed or established by comparative linguistics.

The first group, comprising Tibetan and Chinese dialects, consists of languages whose respective genetic relatedness is a matter of oral and written memory, hence the designation “dialects”, even though both Tibetan and Chinese dialects are, respectively, synchronically as diverse as the Romance or Germanic languages within the Indo-European language family. The knowledge of their respective genetic relatedness and even the accepted division into distinct groups by and large derive from non-linguistic sources: self-awareness, culture, history, and geographical distribution of the groups in question. In this context, existing linguistic classifications have mostly sought to translate this established division into linguistic terms. Chinese and Tibetan dialectal studies commonly hold that most substantial differences between dialects lie in phonology and lexicon, whereas syntactic differences are relatively negligible. Hence, the favoured criteria for classification are for the most part phonological. Several classifications have been advanced for both Tibetan and Chinese dialects (the mostly widely accepted classifications are considered in some detail below).

The second group, Qiangic, brings together languages whose genetic relatedness is not known, but hypothesized.

In terms of a genetic argument, all considered cases (Tibetic, Sinitic, Qiangic) are subgroups. That is to say, they are known or hypothesized monophyletic groups. A subgroup admits as evidence only shared innovations as distinguished from retentions or parallel developments (e.g. Harrison 2003:232-239).

The most widespread classification of Tibetan dialects spoken in China divides all dialects into three groups: Dbus-gtsang (or Central), Khams, and Amdo, which correspond to the three namesake historical provinces of Tibet.⁶ In linguistic terms, this classification is based, among others, on the presence of tone, on the simplification of consonant clusters, on the presence of voiced obstruent initials, on the development of Old Tibetan clusters with medials *-w-*, *-y-* and *-r-*, on the number of consonant codas, and on the presence and the number of diphthongs.⁷ (An example of one classification (Hu 1991) is provided in Figure 1. The three groups of dialects spoken in China are indicated in bold.)

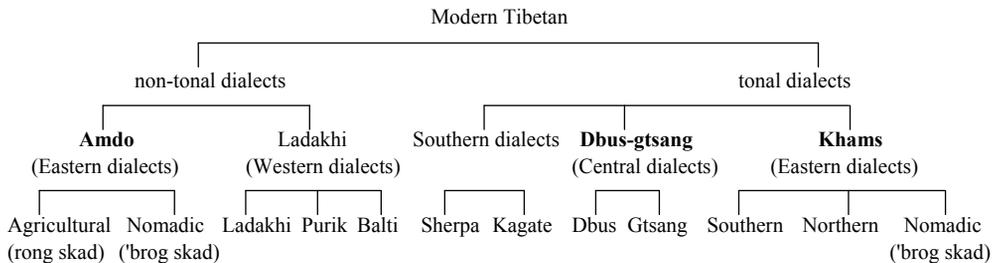


Figure 1: Classification of Tibetan dialects (adapted from Hu 1991:177)

Viewed from the principles of classification outlined above, this classification yields little practical value. Overall, the underlying classification criteria are neither mutually exclusive nor representative for each respective group. As a result, they fail to single out homogeneous groups and to do justice to the actual diversity of Tibetan dialects. Consequently, the resulting classification does not allow for generalizations about the formed groups and it has no predictive power with respect to the assignment of newly

⁶ Cf. Gesang (1964), Qu & Jin (1981), Jin (1983:114-145), Hu (1991:173-177), Zhang (1993, 1996).

⁷ Alternative classifications of Tibetan dialects spoken in China also exist. Targeting the presence of tone as the sole criterion, Hu Tan (1991) arrives at a division of all dialects into just two groups: Khams-gtsang and Amdo (Figure 1). Increasing the number of criteria to five (presence of tone, voiced obstruent initials, consonant clusters, consonant codas and diphthongs), Huang et al. (1994) arrive at four groups: Dbus-gtsang, Khams, Amdo and Yushu 玉樹 (spoken in Zado 雜多 *rdza rdo*, Jiegu 結古 *skye rgu*, Nangqian 囊謙 *nang chen* Counties of Yushu Prefecture in Qinghai Province, and Baqing 巴青 *sbra chen*, Dingqing 丁青 *steng chen* and Suo 索 Counties in Tibet).

described varieties. Take, for instance, the Baima language (also known as the Baima Tibetan dialect) of Northern Sichuan and Southern Gansu provinces in the People's Republic of China (PRC). The linguistic affiliation of Baima has long been disputed, due to the popular belief among the Baima people that they are descendants of the Di people, an ancient Tibeto-Burman ethnic group who inhabited roughly the same area until their gradual assimilation into the Han and the Tibetans during the Tang Dynasty (Chirkova 2008c). If judged solely on linguistic grounds, the language of the Baima people is predominantly Tibetan-like in all its linguistic sub-systems, even though it exhibits some non-Tibetan features in its lexicon, morphology, and syntax (Chirkova 2008b). Accordingly, Baima is argued to be a Tibetan dialect (Huang & Zhang 1995:104, Zhang 1997:134-135, 140). The problem, however, is that this new Tibetan dialect cannot be straightforwardly accommodated by the dominant tripartite dialect classification of Tibetan dialects spoken in China.⁸ Based on the exhibited characteristics, Baima is at the same time Khams-like (given the drastic simplification of the syllable structure, the elimination of Old Tibetan syllable codas, and the presence of tones) and Amdo-like (given, for instance, the development of Old Tibetan velar stop clusters with medial *-r-* into alveopalatal affricates and of Old Tibetan dental and labial stop clusters with the same medial into retroflex affricates).⁹ In sum, the dominant classification of Tibetan dialects of China fails to accommodate Baima. In fact, the same holds true for any of the newly discovered dialects in its neighbourhood, such as Zhongu (J. Sun 2003a), Chos-rje (J. Sun 2003b), or the Tibetan dialects of Western Sichuan (Suzuki 2008).

In his discussion of the popular tripartite classification of the Tibetan dialects of China, Jackson T.-S. Sun (2003a:794-797) points out that this classification mixes possibly convergent, but independent developments (such as the presence of tone) with shared retentions (such as the preservation of Old Tibetan voiced obstruent onsets). This, in Sun's analysis, explains its inadequacy. The envisaged solution is to upgrade this routinely accepted (but overall unrevealing) classification into a meaningful (i.e. genetic) classification, which, in turn, should be based on shared innovations and phonological isoglosses to delimit precise dialect boundaries. Sun does not discuss the feasibility of establishing a classification along these lines and, to my knowledge, such a classification has not yet been attempted. However, a close investigation of a possibility of such a *stricto sensu* genetic classification of Chinese dialects by Laurent Sagart (1998) suggests that such a classification is not viable.

The currently widely accepted classification of Chinese dialects is based on one single criterion: the development of the Middle Chinese voiced stops into the modern

⁸ Presumably for this reason, Baima is not included in any of the comprehensive outlines of Tibetan dialects of China (e.g. Zhang 1996).

⁹ Chirkova (2008c, 2008d).

dialects (Li 1937, Yuan et al. 2001).¹⁰ While this criterion works fairly well in the majority of cases, as Norman (1988:181) points out, it is clearly valid only for a relatively recent historical period, given that the voiced initials probably persisted in the North down to at least the tenth or eleventh centuries. This traditional classification of Chinese dialects suffers from the same flaws as its Tibetan counterpart. It fails to do justice to the actual diversity of dialects and to form a reliable basis for generalizations regarding the formed groups. This problem is particularly acute in relation to dialect syntax (e.g. Chirkova 2008a for the Mandarin group).

The insufficiency of one criterion for a meaningful classification of Chinese dialects in general, and Hakka dialects in particular, has been pointed out by Sagart (1998), who further notes that a choice of some other criterion would result in an entirely different classification.

Similar to J. Sun's assessment of the current classifications of Tibetic languages, Sagart views the non-conformity of the traditional classification of Chinese dialects to the standards of genetic classification as one reason for its shortcomings. He notes that the traditional classification of Chinese dialects defines dialect zones whose historical statuses are very different: some are defined by innovation, others by retention, whereas it is generally admitted that only innovations are useful in linguistic classification. However, in stark contrast with J. Sun's conclusion, Sagart finds a *stricto sensu* genetic classification of Chinese dialects impossible and even unnecessary. It is impossible, Sagart argues (1998:299), because innovations may be obliterated or reversed through contact. Hence, dialect groups that are currently recognized in Chinese dialectology are fuzzy entities that owe much of their make-up to contact (i.e. horizontal transmission) as opposed to vertical inheritance.¹¹ This is not to say that a genetic classification of Chinese dialects is altogether unattainable. The linguistic history of the group can be reconstructed to the extent possible from written sources. However, as Sagart argues, a genetic dialect classification is unnecessary, because such a classification has at best a minor role to play in the furthering of knowledge on Chinese dialects. Instead, Sagart points out, geographical approaches to Chinese dialectology hold more promise.

The use of a small number of criteria and the use of common innovations as a valid criterion are disfavoured as basis for a meaningful classification in the case of well-documented dialects or languages *known* to be related (Sinitic). Curiously, these criteria gain in explanatory value in the case of little-known languages that are *not known* to be

¹⁰ Other classifications of Chinese dialects exist, based on different sets of criteria, e.g. Norman (1988) and Wurm et al. (1988). These classifications by and large correspond to the currently widely accepted division of Chinese dialects into groups in the tradition classification by Li (1937).

¹¹ This is a received view in Chinese dialectology, e.g. Pulleyblank (1991, 1998).

related. Take the Qiangic languages of the Chinese South-West as an example.

Qiangic is a putative group of eleven Tibeto-Burman languages of Sichuan and Yunnan provinces in the PRC. After H. Sun (1983, 2001), these eleven languages are often grouped in a northern branch and a southern branch.¹² The Qiangic subgroup in its current understanding was proposed in the 1970s by Chinese linguists.¹³ It brings together geographically adjacent, non-written, and under-researched languages, which are not known to be related. This subgroup relies primarily on shared typology: the widespread use of measure words, directional prefixes and similar case marking, to which some, equally general phonological features such as the presence of uvulars and tones have also been added (H. Sun 1983). Seen from mainstream historical linguistics, these are type-identifying features (Nichols 1996:48), in other words, features that are found in enough unrelated language families and that are consequently low in identificational value as to the genetic relationship. Overall, the percentage of common vocabulary among the languages of the Qiangic subgroup (the feature which, in contrast to the typological characteristics above, could provide more reliable support for the hypothesis of a common origin of the languages) is relatively small. Altogether it makes up less than 20% between any two Qiangic languages (Huang 1991:355). Notably, the type-identifying features in Qiangic are not found in their totality in languages from the neighboring subgroups. At the same time, languages of these neighboring (and phylogenetically better-understood) subgroups spoken in the Qiangic area, appear to share with Qiangic languages many features that are held specific of the Qiangic subgroup. An example of such a language is Kami, a dialect of Tibetan spoken in the southern end of the Qiangic area. Kami has typically Qiangic and at the same time strikingly non-Tibetan features, such as directional prefixes, measure words, or uvular phonemes (Chirkova forthcoming). The presence of these ‘Qiangic’ features in a Tibetic language suggests that Qiangic is rather a *Sprachbund*: a geographically delimited area that includes languages that are genetically unrelated, but share a fair number of highly distinctive traits.

Even though the grounds for its classification are demonstrably weak, the Qiangic subgroup has found its way in all present classifications of the Tibeto-Burman language family. This is perhaps because even despite their diversity, Qiangic languages are possibly more closely related to each other than to anything else in Tibeto-Burman (Bradley 1997:36).

¹² Northern Qiangic languages include Qiang, rGyalrong (Jiarong), Ergong, Prinmi (Pumi) and Minyak (Muya). Southern Qiangic languages comprise Zhaba, Choyo (Queyu), Ersu, Namuyi, Xumi (a.k.a. Shixing) and Guiqiong.

¹³ For the history of the Qiangic subgroup, see J. Sun (1992).

Formulated on such loose grounds, the Qiangic subgroup fails the test of a meaningful classification: it cannot serve as basis for generalizations or single out features specific for the group. In fact, the precise composition of this subgroup has been debated since its very proposal. In the 1980s, the aforementioned Baima language was argued to be Qiangic, because it exhibits some features that are held to be typical of this subgroup, such as the presence of directional prefixes (H. Sun 1980). Conversely, in the 1990s, one of the Qiangic languages proper, Namuyi, was claimed to be genetically related to the Ngwi languages instead (Lama 1994, Huang 1997:13-15), based on a large amount of basic vocabulary shared between Ngwi and Namuyi.

Recent years have witnessed an upsurge of interest in Qiangic languages and linguistics. Many Qiangic languages are currently being investigated. Attempts are also undertaken to place this methodologically problematic grouping on a scientific footing: by focusing on common innovations, lexical comparisons and, eventually, phonological reconstructions.¹⁴ Viewed in the context of the above discussion, such a classification is bound to be based on few criteria, which are furthermore spurious, as no objective criteria to distinguish common features (in lexicon or elsewhere) from accidental similarity in form, borrowing, and/or genetic drift have yet been found (cf. Thomason & Kaufman 1988:6).¹⁵

In fact, there is not necessarily a correlation between the number of underlying criteria and the quality of the resulting classification. Even one character may be sufficient, if it is known to be correlated with many others of varied nature. If the correlation has been established, the use of the one feature will amount to the use of all correlated features together. Unfortunately, to my knowledge, lexical sharings are not correlated with any other linguistic feature. The use of this one single criterion is hence bound to result in a biased classification, whose validity is furthermore contingent on the acceptance of lexical sharings as probative of genetic relatedness.

What would be the gains of the genetic approach in the case of Qiangic languages, if pursued rigorously? A genetic argument in the subgroup context can ultimately lead to either of these two possibilities:

¹⁴ The establishment of a genetic relationship in the Tibeto-Burman context must rely—*faute de mieux*—on shared derivational morphology and lexical comparisons, in spite of the fact that both can be borrowed, given a sufficient degree of contact. Because few inflectional paradigms can be identified across the family, shared and especially irregular inflectional paradigms cannot be relied on to prove the existence of the family (Handel 2008:435).

¹⁵ Subgrouping is known in mainstream comparative linguistics to be methodologically and factually problematic. However, its limitations are argued away on the basis of the practical necessity of subgrouping assumptions for the application of the comparative method (Harrison 2003:239).

First, the comparative method could show that the Qiangic subgroup is polyphyletic. The fact that Qiangic languages have grown to share a number of salient features is then to be explained through effects of convergence and hybridisation. Notably, the same conclusion can be directly arrived at by a closer investigation of empirical data, without resorting to the formal apparatus of lexical comparisons, loanword stratification, search for cognate etyma, establishment of sound correspondences, and phonological reconstructions.

Second, the comparative method could show that the Qiangic group is monophyletic. In that case, the extreme synchronic diversity of the languages forming the group, and the relatively meagre amount of shared cognate vocabulary, would have to be explained. The comparative method cannot provide this explanation and does not even posit such among its explicit goals. Consequently, a closer investigation of empirical data will again be required if any sense is to be made of the linguistic situation and the historical development of the relevant languages.

The following three conclusions can be made on the basis of the three cases reviewed in this section.

First, the dominant status of genetic classification and the expectations associated therewith are at odds with the actual practical merits of genetic classification. After all, genetic classification reflects known or hypothesized lines of descent, and in so doing, it restricts the available data in a way that excludes much information.

Second, in those cases where the line of descent is unknown (e.g. Qiangic), to proceed with an explanatory approach (genetic classification) before assembling a representative set of empirical data, runs the risk of applying potentially subjective criteria and of involving circular reasoning.

Third, as suggested by studies of well-documented languages that are known to be related (Sinitic, Tibetic), it is impossible to combine information on genetic descent with maximum information on synchronic similarities between the classified varieties into one classification scheme. Therefore, at least two approaches (and classifications) are warranted to organize these two types of information, for example, one genetic and one geographical (cf. Sagart for Sinitic).

Overall, given the little practical value of existing classifications of Tibeto-Burman languages, genetic or otherwise, and given the overall pressure to reflect on phylogenies, a linguist working with Tibeto-Burman languages cannot but get involved in classification work. Indeed, the lion's share of production in Tibeto-Burman linguistics deals with genetic classification issues: from individual languages to subgroups, subgroups to families, and families to macro-families (cf. Handel 2008). While existing linguistic classifications are far from satisfactory and no immediate linguistic solutions are in sight, it may be worthwhile, given the success of the "language as an organism" metaphor, to

turn to biology for inspiration and possible solutions.

4. From classifications in biology to classifications in linguistics

In its simplest form, classification in biology is the ordering of entities into groups on the basis of their relationships, that is, of their associations by contiguity, similarity, or both (cf. Simpson 1961:9, quoted from Sokal & Sneath 1963:3).¹⁶ Notably, the term “relationship” may imply relationship by ancestry or, alternatively, it may indicate the overall similarity (or affinity) as judged by the characters of the organisms without any implication as to their relationship by ancestry. It has been argued that no classificatory scheme is able simultaneously to yield information on the degree of synchronic similarity and genetic descent (Michener & Sokal 1957, Sokal & Sneath 1963:55-57). Consequently, two major types of classification exist: empirical (synchronic) and genetic (diachronic). The empirical classification is often referred to as “natural”, i.e. denoting a system that occurs naturally in the empirical world. The two types of classification, natural and genetic, differ in scope, in purpose, and in the number of underlying criteria.

A natural classification is one that endeavors to utilize many attributes of the classified entities and to accurately reflect most of the various natures of the objects. While it is not explanatory as to the biological phenomena that it describes, a natural classification forms solid basis for generalizations concerning the described data and it is useful for a wide range of purposes (e.g. Gilmour 1940:472). An example of an empirical classification in biology, that groups animals and kinds according to their shared physical characteristics, is the familiar Linnaean taxonomy widely used in the biological sciences.¹⁷

A natural system of classification in biology is contrasted with more or less ‘artificial’ systems, which have no natural occurrence prior to their construction by the researcher. Such an ‘artificial’ classification is based on a limited range of attributes, selected with an eye to a specific classification purpose. Consequently, it is useful for limited purposes (Blackwelder 1967:186).

Given that evolution is responsible for the existence and structure of the natural system, classification in biology (as in linguistics) is possible because evolution has

¹⁶ The compilation in this section is based on Huxley (1940), Sokal & Sneath (1963), Blackwelder (1967), Jardine & Sibson (1971), Salthe (1972), Mayr (1976, 1988), Bailey (1994), and Pigliucci & Kaplan (2006).

¹⁷ A phonetic view of classification in biology, according to which the sole aim of classifications is to reflect as accurately as possible the relative similarities or dissimilarities of populations, is frequently said to be pioneered by Michel Adanson in his *Familles des Plantes* (1763) (cf. Jardine & Sibson 1971:136-137).

produced diverse kinds (or languages). From this it is possible to generalize that classification is based on evolution, but it must be recognized that this means that it is based on the results of evolution: empirical facts, synchronically observable concrete living organisms. Evolution can naturally be seen as the central cause in biology. Nonetheless, it cannot be used as an explanatory concept or an independent basis for classification. The main reason for that is that common ancestry is never known as a fact, but merely hypothesized. Furthermore, the phylogeny is based, on the one hand, on previous classifications and, on the other hand, on known empirical facts, so that there are no examples of classifications that are actually based on what were taken to be phylogenetic facts as distinct from the features of living organisms (Sokal & Sneath 1963:56-57, Blackwelder 1967:191-193). The genetic type of classification, which focuses on providing a summary of existing knowledge of phylogeny, is hence a subsidiary classification (e.g. Gilmour 1940:473).

To return to the field of linguistics, we can observe that linguistics does not have in its arsenal any analogy to natural classification in biology. Instead, genetic classification has come to be used as an explanatory concept and a general linguistic classification. It is, however, a secondary classification, because it selects few criteria (essentially basic vocabulary in the domain of Tibeto-Burman languages) and it pursues one specific domain of inquiry (the common ancestry of languages). As such, it falls short of accounting for the versatility of empirical data, and it is unable to accommodate new data, as discussed above on the basis of Tibetic, Sinitic, and Qiangic languages. While a significant and intriguing facet of linguistic inquiry, indispensable for the purposes of historical linguistics, it is unsuitable as general linguistic classification.

Successful in those cases where written records (linguistic “fossil record”) are available and linguistic similarities self-evident (through history or mutual intelligibility), genetic classification reaches its limits in those cases where relationships between languages are less evident and written records are scarce or absent. Since no reliable method to ascertain characters probative of historical relatedness has been developed, much caution is required when using genetic arguments for classification purposes. Of course, there is no harm in speculating about possible relationships between languages in the latter type of situation, but in linguistics, as in biology (cf. Sokal & Sneath 1963:8), it appears more informative to concentrate on collection and analysis of the maximum of empirical facts, because facts inform and advance progress in linguistic theory and historical linguistics.

5. Discussion

To return to my departing point as a practitioner in need of a practical linguistic

classification, I cannot but conclude that a genetic classification might not be “the most satisfying way to group languages” for “professional linguists and general readers”, as it is praised to be.

There is need for another general linguistic classification, which will be satisfying in the sense that it can give justice to the versatility of the subject matter of linguistics, while providing a practical means to organize our knowledge of various facets of the continuously growing body of data. To serve this purpose, this classification is bound to be empirical, i.e. based on a variety and quantity of features.

On a more abstract level, in classification as elsewhere in linguistics, there is no fundamental contrast between description and explanation, as there is naturally need for both descriptive and explanatory approaches (Dryer 2006). While there is no intrinsic incompatibility in describing a language and then explaining the things described, it appears desirable that both types of approaches enjoy equal footing, so that explanatory approaches do not channel descriptive ones. The primary goal in linguistics, as in any area of scientific inquiry, remains first and foremost to describe a set of facts without any particular theoretical implication. This set of facts is further open to synthesis, hypothesis generation and interpretation, including, for instance, grading of observed similarities in terms of their respective historical status (retentions, innovations, parallel developments), and assessing the balance between horizontal and vertical lines of transmission in language change. All in all, my claim is that we need both empirical and genetic classifications, as well as many other types of classification (for instance, typological) for various specific purposes.

In terms of close parallels between linguistics and biology, the two disciplines continue to go hand in hand. Both owe their success to the theory of evolution with the tree model as its central concept. This theory and the tree concept have all along exerted strong influence on the development of the two disciplines. For over a century, linguistics and biology were mainly concerned with fleshing out a universal tree in their respective disciplines. The ongoing revision of this concept in biology, where evidence begins to mount up that a substantial component of the evolutionary process is non-tree-like (Doolittle 1999, Dagan & Martin 2006, Lawton 2009), is bound to have a major impact on linguistics. In fact, linguistics (where the need to incorporate horizontal transmission into ideas about linguistic evolution has long been evident) is already undergoing a similar paradigm shift.

In sum, further progress is possible only if a broad range of varied data is taken into consideration, and this broad range of data requires well-grounded theories and adequate classifications to be well-understood.

References

- Aikhenvald, Alexandra Y., and R. M. W. Dixon. (eds.) 2001. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford & New York: Oxford University Press.
- Adanson, Michel. 1763. *Familles des Plantes*. Paris: Vincent.
- Bailey, Kenneth D. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Thousand Oaks, London & New Delhi: SAGE.
- Benedict, Paul K. 1972. *Sino-Tibetan: A Conspectus*. Cambridge: Cambridge University Press.
- Blackwelder, Richard E. 1967. *Taxonomy: A Text and Reference Book*. New York & London: John Wiley & Sons, Inc.
- Blake, Barry J. 2006. Classification of languages. *Encyclopedia of Languages and Linguistics*, ed. by Keith Brown et al., 446-457. Amsterdam: Elsevier.
- Bradley, David. 1997. Tibeto-Burman languages and their classification. *Tibeto-Burman Languages of the Himalayas*, ed. by David Bradley, 1-72. Canberra: The Australian National University.
- Campbell, Lyle, and William J. Poser. 2008. *Language Classification: History and Method*. Cambridge & New York: Cambridge University Press.
- Chappell, Hilary. 2001. Contact and diffusion in Sinitic. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, ed. by Alexandra Y. Aikhenvald & R. M. W. Dixon, 328-357. Oxford & New York: Oxford University Press.
- Chirkova, Katia. 2008a. *Gěi* 'give' in Beijing Mandarin. *Cahiers de Linguistique – Asie Orientale* 37.1:3-42.
- Chirkova, Katia. 2008b. On the position of Baima within Tibetan: a look from basic vocabulary. *Evidence and Counter-evidence: Essays in Honour of Frederik Kortlandt*, Vol. 2, ed. by Alexander Lubotsky, Jos Schaeken & Jeroen Wiedenhof, 69-91. Amsterdam & New York: Rodopi.
- Chirkova, Katia. 2008c. Baima Zangzu wei Dizu shuo zhiyi [The Baima Tibetans and the Di people of Chinese historical records: challenging the link]. *Bulletin of Chinese Linguistics* 3.1:167-180.
- Chirkova, Katia. 2008d. Baimayu shizheng fanchou ji qi yu Zangyu fangyan de bijiao [Evidentials in Baima and Tibetan dialects compared]. *Minzu Yuwen [Minority Languages of China]* 2008.3:36-43.
- Chirkova, Katia. (forthcoming). Phonological profile of Kami, the Tibetan dialect of Muli. Special issue of Language and Linguistics on lesser known dialects of Tibetan, ed. by Jackson T.-S. Sun.

- Curnow, Timothy Jowan. 2001. What languages features can be 'borrowed'? *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, ed. by Alexandra Y. Aikhenvald & R. M. W. Dixon, 412-436. Oxford & New York: Oxford University Press.
- Dagan, Tal, and William Martin. 2006. The tree of one percent. *Genome Biology* 7.10:118.1-118.7.
- Dench, Alan. 2001. Descent and diffusion: the complexity of the Pilbara situation. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, ed. by Alexandra Y. Aikhenvald & R. M. W. Dixon, 105-133. Oxford & New York: Oxford University Press.
- Dixon, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge & New York: Cambridge University Press.
- Doolittle, W. Ford. 1999. Phylogenetic classification and the universal tree. *Science* 284: 2124-2128.
- Driem, George van. 2004. Language as organism: a brief introduction to the Leiden theory of language evolution. *Studies on Sino-Tibetan Languages: Papers in Honor of Professor Hwang-chenng Gong on His Seventieth Birthday*, ed. by Ying-chin Lin, Fang-min Hsu, Chun-chih Lee, Jackson T.-S. Sun, Hsiu-fang Yang & Dah-an Ho, 1-9. Taipei: Institute of Linguistics, Academia Sinica.
- Driem, George van. 2008. The language organism: parasite or mutualist? *Evidence and Counter-evidence: Essays in Honour of Frederik Kortlandt*, Vol. 2, ed. by Alexander Lubostky, Jos Schaeken & Jeroen Wiendehof, 101-112. Amsterdam & New York: Rodopi.
- Dryer, Matthew. 2006. Descriptive theories, explanatory theories, and basic linguistic theory. *Catching Language: Issues in Grammar Writing*, ed. by Felix Ameka, Alan Dench & Nicholas Evans, 207-234. Berlin & New York: Mouton de Gruyter.
- Gesang, Jumian (skal bzang 'gyur med). 1964. *Zangyu Fangyan Gaiyao / Bod kyi yul skad gces bsdus [Essentials of Tibetan Dialects]*. Beijing: Minzu University of China.
- Gilmour, J. S. L. 1940. Taxonomy and philosophy. *The New Systematics*, ed. by Julian Huxley, 461-474. Oxford: Oxford University Press.
- Handel, Zev. 2008. What is Sino-Tibetan? Snapshot of a field and a language family in flux. *Language and Linguistics Compass* 2.3:422-441.
- Harrison, Shelly P. 2003. On the limits of the comparative method. *The Handbook of Historical Linguistics*, ed. by Brian D. Joseph & Richard D. Janda, 213-243. Malden: Blackwell.
- Hu, Tan. 1991. Zangyu [The Tibetan language]. *Han-Zangyu Gailun [An Introduction to Sino-Tibetan Languages]*, Vol. 2, ed. by Xueliang Ma, 109-207. Beijing: Peking University Press.

- Huang, Bufan. 1991. Qiangyuzhi [Qiangic languages]. *Han-Zangyu Gailun [An Introduction to Sino-Tibetan Languages]*, Vol. 2, ed. by Xueliang Ma, 208-369. Beijing: Peking University Press.
- Huang, Bufan. 1997. Tongyuanci bijiao cibiao de xuanci fanwei he biao zhun—yi Zang-Mianyu tongyuanci bijiao cibiao de zhiding wei li [On the scope and criteria of word selection for comparative cognacy lists, exemplified by a draft list of cognate words in Tibeto-Burman languages]. *Minzu Yuwen [Minority Languages of China]* 1997.4:10-16.
- Huang, Bufan, Suonan Jiangcai, and Minghui Zhang. 1994. Yushu Zangyu de yuyin tedian he lishi yanbian guilü [Characteristics in Yushu Tibetan phonology and its rules of historical change]. *Tibetology in China* 26.2:111-134.
- Huang, Bufan, and Minghui Zhang. 1995. Baimahua zhishu wenti yanjiu [A study of the genetic affiliation of Baima]. *Tibetology in China* 1995.2:79-118.
- Huxley, Julian. (ed.) 1940. *The New Systematics*. Oxford: Oxford University Press.
- Jardine, Nicholas, and Robin Sibson. 1971. *Mathematical Taxonomy*. London: John Wiley & Sons, Ltd.
- Jin, Peng. 1983. *Zangyu Jianzhi [Outline of the Tibetan Language]*. Beijing: Nationalities Press.
- Janda, Richard D., and Brian D. Joseph. 2003. On language, change, and language change — or, of history, linguistics, and historical linguistics. *The Handbook of Historical Linguistics*, ed. by Brian D. Joseph & Richard D. Janda, 3-180. Malden: Blackwell.
- Kortlandt, Frederik H. H. 1985. A parasitological view of non-constructive sets. *Studia linguistica diachronica et synchronica: Werner Winter, sexagenario anno MCMLXXXIII*, ed. by Ursula Pieper & Gerhard Stickel, 477-483. Berlin: Mouton de Gruyter.
- Kortlandt, Frederik H. H. 2003. The origin and nature of the linguistic parasite. *Language in Time and Space: A Festschrift for Werner Winter on the Occasion of His 80th Birthday*, ed. by Brigitte L. M. Bauer & Georges-Jean Pinault, 241-244. Berlin & New York: Mouton de Gruyter.
- Lama, Ziwo. 1994. Namuyiyu zhishu yanjiu [On the genetic affiliation of the Namuyi language]. *Minzu Yuwen [Minority Languages of China]* 1994.1:50-60.
- Lawton, Graham. 2009. Why Darwin was wrong about the tree of life. *New Scientist* 2692:34-39.
- Li, Fang-kuei. 1937. Chinese languages and dialects. *The Chinese Yearbook 1937*, 59-65. Shanghai: The Commercial Press. Reprinted in *Journal of Chinese Linguistics* 1.1:1-13, 1973.

- Matisoff, James A. 2001. Genetic versus contact relationship: prosodic diffusibility in South-East Asian languages. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, ed. by Alexandra Y. Aikhenvald & R. M. W. Dixon, 291-327. Oxford & New York: Oxford University Press.
- Mayr, Ernst. 1976. *Evolution and the Diversity of Life: Selected Essays*. Cambridge: Belknap Press of Harvard University Press.
- Mayr, Ernst. 1988. *Toward a New Philosophy of Biology: Observations of an Evolutionist*. Cambridge: Belknap Press of Harvard University Press.
- Michener, Charles D., and Robert R. Sokal. 1957. A quantitative approach to a problem in classification. *Evolution* 11:130-162.
- Mufwene, Salikoko S. 2001. *The Ecology of Language Evolution*. Cambridge & New York: Cambridge University Press.
- Nichols, Johanna. 1996. The comparative method as heuristic. *The Comparative Method Revisited: Regularity and Irregularity in Language Change*, ed. by Mark Durie & Malcolm Ross, 39-71. Oxford & New York: Oxford University Press.
- Noonan, Michael. 2008. Genetic classification and language contact. Heidelberg: Bibliothek des Südasien-Instituts der Universität Heidelberg. Online-Ressource. Electronic publications by Michael Noonan; 23: <http://archiv.ub.uni-heidelberg.de/savifadok/volltexte/2008/215>.
- Norman, Jerry. 1988. *Chinese*. Cambridge & New York: Cambridge University Press.
- Peyraube, Alain. 2005. L'homme aux six mille langues. *Aux origines des langues et du langage*, ed. by Jean-Marie Hombert, 450-461. Paris: Editions Fayard.
- Peyraube, Alain. 2007. Languages and genes in China and in East Asia. *Bulletin of Chinese Linguistics* 2.1:1-15.
- Pigliucci, Massimo, and Jonathan Kaplan. 2006. *Making Sense of Evolution: The Conceptual Foundations of Evolutionary Biology*. Chicago: University of Chicago Press.
- Popper, Karl R. 1969. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge & Kegan Paul. First published in 1963.
- Pulleyblank, Edwin G. 1991. Chinese dialect studies. *Languages and Dialects of China*, ed. by William S-Y. Wang, 431-456. Berkeley: Project on Linguistic Analysis.
- Pulleyblank, Edwin G. 1998. *Qieyun* and *Yunjing*: the essential foundation for Chinese historical linguistics. *The Journal of the American Oriental Society* 118.2:200-216.
- Qu, Aitang, and Xiaojing Jin. 1981. Zangyu fangyan de yanjiu fangfa [Methods in the study of the Tibetan dialects]. *Journal of Southwest University for Nationalities* 1981.3:76-83.
- Ruhlen, Merritt. 1994. *On the Origins of Languages: Studies in Linguistic Taxonomy*. Stanford: Stanford University Press.

- Sagart, Laurent. 1998. On distinguishing Hakka and non-Hakka dialects. *Journal of Chinese Linguistics* 26.2:281-302.
- Salthe, Stanley N. 1972. *Evolutionary Biology*. New York: Holt, Rinehart and Winston.
- Simpson, George Gaylord. 1961. *Principles of Animal Taxonomy*. New York: Columbia University Press.
- Sokal, Robert R., and Peter H. A. Sneath. 1963. *Principles of Numerical Taxonomy*. San Francisco: W. H. Freeman & Company.
- Sun, Hongkai. 1980. Baimaren de yuyan [The language of the Baima]. *Baima Zangren Zushu Wenti Taolunji [A Collection of Discussions on the Question of the Ethnic Classification of the Baima Tibetans]*, ed. by Sichuan Nationalities Research Institute, 15-26. Chengdu: Sichuan Nationalities Research Institute.
- Sun, Hongkai. 1983. Liujiang liuyu de minzu yuyan ji qi xishu fenlei [Minority languages of the six river valley and their genetic classification]. *Minzu Xuebao [Scholarly Journal of Nationalities]* 3:99-273.
- Sun, Hongkai. 2001. Lun Zang-Mian yuzu zhong de Qiangyuzhi yuyan [On the Qiangic branch of the Tibet-Burman language family]. *Language and Linguistics* 2.1:157-181.
- Sun, Jackson T.-S. 1992. Review of *Zang-Mianyu Yuyin he Cihui [Tibeto-Burman Phonology and Lexicon]*. *Linguistics of the Tibeto-Burman Area* 15.2:73-113.
- Sun, Jackson T.-S. 2003a. Phonological profile of Zhongu: a new Tibetan dialect of Northern Sichuan. *Language and Linguistics* 4.4:769-836.
- Sun, Jackson T.-S. 2003b. Qiuji Zangyu de yuyin tezheng [Phonological profile of Chos-rje Tibetan]. *Minzu Yuwen [Minority Languages of China]* 2003.6:1-6.
- Suzuki, Hiroyuki. 2008. Nouveau regard sur les dialectes tibétains à l'est d'Aba: Phonétique et classification du dialecte de Shangkhog [Songpan-Jiuzhaigou]. *Linguistics of the Tibeto-Burman Area* 31.1:85-108.
- Thomason, Sarah Grey. 2000. On the unpredictability of contact effects. *Estudios de Sociolinguística* 1.1:173-182.
- Thomason, Sarah Grey, and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Wurm, S. A. et al. 1988. *Language Atlas of China*. Hong Kong: Longman.
- Yuan, Jiahua et al. 2001. *Hanyu Fangyan Gaiyao [Outline of Chinese Dialects]*. Beijing: Language and Culture Press.
- Zhang, Jichuan. 1993. Zangyu fangyan fenlei guanjian [Opinions on the classification of Tibetan dialects]. *Minzu Yuwen Lunwenji [Papers on Minority Languages]*, ed. by Qingxia Dai, Meizhen Luo & Yingxin Yang, 297-309. Beijing: Minzu University of China Press.

Katia Chirkova

Zhang, Jichuan. 1996. A sketch of Tibetan dialectology in China: classifications of Tibetan dialects. *Cahiers de Linguistique – Asie Orientale* 25.2:113-144.

Zhang, Jichuan. 1997. Particularités phonétiques du Baima. *Cahiers de Linguistique – Asie Orientale* 26.1:131-153.