Diversity in Sinitic Languages

Edited by
HILARY M. CHAPPELL

OXFORD UNIVERSITY PRESS
Introduction: Ways of tackling diversity in Sinitic languages

HILARY M. CHAPPELL

This volume represents one of the first major outcomes of the ERC SINOTYPE project (2009–2013) which brought together seven researchers from China, France, and Australia to carry out empirically-based typological studies on the vast, yet little explored domain of Sinitic languages.1

Challenging the tradition in Chinese linguistics, the SINOTYPE project has consistently treated Chinese dialects as a separate, though unquestionably affiliated, group of languages belonging to the Sino-Tibetan language family and having the technical name of ‘Sinitic’. They are generally considered to be a first-order, sister branch to the Tibeto-Burman languages, the latter lying to the west and southwest of the core area in China where Sinitic languages are found. Details of their classification and a linguistic map are presented in Chapter 5 for the ten recognized second-order branches of Sinitic: Mandarin, Jin, Wu, Hui, Gan, Xiang, Yue, Hakka, Min, and Pinghua.

The issue of linguistic diversity, genetic relationship, and areal linguistics has been of fundamental importance to the SINOTYPE project, since any typological classification that involves a set of parameters applied to a number of languages located in contiguous zones, related or not, will have a tendency to transcend genetic relationships and, in the case of Sinitic, to be positively oriented towards the setting up of linguistic areas within China, grouping together languages and dialects from different branches of Sinitic. Admittedly, in this instance, we are treating related languages and dialects and, thus, cases of apparent ‘interdialectal’ language contact and diffusion, yet the same principles relevant for areal linguistics in general (Clark 1989; Dahl 2001; Enfield 2005) and grammaticalization zones (Heine and Kuteva 2005, 2006) can be seamlessly applied in such cases, as Chapter 2 will argue. The relevant connections between genes, family tree relationships, languages, and diffusibility

1 A description of the SINOTYPE project is given in the preface to this volume.
are treated in both Matisoff (2001) and Peyraube (2007) for Southeast Asia and East Asia respectively.

In reality, the Sinitic languages are, *grosso modo*, as distinct from one another as European languages, a point argued in Chappell (2001a). Even within the second-order groups of Sinitic, such as the large Mandarin dialect area with its estimated 800 million L speakers today, a further division into eight main subgroups is necessary to account for its diversity—and this merely in the form of differences in their phonological inventories, tone systems, and lexicon, to which the *Language Atlas of China* amply bears testimony (Wurm et al. 1987, 2nd edition 2013). The work on syntactic differences is only beginning.

Moreover, in comparing any two representative dialects for these subgroups of Mandarin, we could quite appropriately discuss issues related to ‘interdialectal’ or ‘intervarietal’ differences and the outcomes of contact, just as we might for the varieties of Irish and Scottish English, or regional dialects such as Yorkshire and Geordie, in relation to Standard British English (SE). Similarly to the case for SE, only a small proportion of the Mandarin population are actually native speakers of the variety closest to Standard Mandarin or *pǔtōnghuà* 普通話, based on the Beijing dialect. In contrast to this, if we compared features of the Puxian isolate of the Min dialect group, located in the southeastern coastal province of Fujian, with the Xi’an dialect of Central Plains Mandarin in Shaanxi, northern China, some 1,800 km away, this would be more akin to a comparison between English and Swedish, or Portuguese and Romanian, and thus truly interlinguistic, despite the evident genetic relationship in all cases.

The closely intertwined connection between typology, areal linguistics, and linguistic diversity has become an indisputable reality today, since the findings of typological classification clearly feed into the identification of linguistic areas and these, in turn, identify the regions where the greatest—or the least—diversity is to be found. Dryer’s detailed studies (2003, 2008) on the typological features of Sino-Tibetan and mainland Southeast Asian languages are a case in point. By way of contrast, this diversity—which is only starting to be measured and appreciated for Sinitic languages—has to be ultimately linked with migrations and consecutive episodes of language contact, particularly in the North-to-South direction in China over the past two millennia (Chappell 2001a and You R. 1992 on migration patterns in China and the formation of Sinitic; LaPolla 2001 on migration patterns and the formation of Sino-Tibetan).

In other words, the explanations for the emergence of linguistic diversity in China have to be precisely formulated in terms of sociolinguistic parameters, including population movements, social behaviour, and social attitudes towards language, that repeatedly arise in language contact situations and condition external borrowing, if not the adoption of another language in part or in whole (Thomason and Kaufman 1988; Mufwene 2001). These considerations of language-external causes leading to diversity undeniably include features otherwise not to be predicted for Sinitic and are principally discussed in Chapter 7 with respect to the area of Far Southern China and
the interaction between Pinghua, Cantonese Yue, and Northern Zhuang (Tai-Kadai). The case is also debated for non-Sinitic languages as the putative source of word order change for ditransitives and adverbial placement in certain dialect groups in Chapter 3, while the possibility of an external model for Sinitic Surpass comparatives is weighed up in Chapter 6. Language contact in Sinitic languages and some of its consequences in the form of stratification, convergence, and hybridization are treated in an earlier article of mine (Chappell 2001b).

The problem of the mixed typology of Sinitic languages is a general one and has been specifically addressed by the SINO TYPE project as a crucial area to target. Hence, one of the objectives has been to find explanations for the contradictory mixture of word order correlations in Sinitic, some of which tally with VO while others tally with OV. The project has thus set out to examine and seek explanations for this apparent hybrid typology found in most Sinitic languages, all of which reveal a perplexing mixture of head-final and head-initial features for languages whose basic word order is SVO.

Consequently, Sinitic languages pose several striking counterexamples to classic Greenbergian word order correlations, in particular, the word order for the main type of comparative construction which is Marker-Standard-Adjective/VP. As is well known, this is the order which correlates with SOV languages (Dryer 1992) and is discussed in Chapter 2, Chapter 6, and Chapter 10. The same disharmonic situation applies for prenominal relative clauses which merely reflect the overall head-final requirement for the noun phrase in Sinitic, yet once again, does not tally with its SVO clause-level word order. Preliminary findings show that, in contrast to Standard Chinese, many Sinitic languages show the use of postnominal relative clauses and head-marked comparatives which evidently harmonize ‘correctly’ with SVO basic word order. Many of these theoretical issues have been separately treated in Chappell, Li, and Peyraube (2007).

Diachronic grammar provides another important methodology for explaining how grammatical change creates linguistic diversity in Sinitic languages. This source, including the language-internal mechanisms of reanalysis, analogy, and exaptation (the recycling of a form into a new unrelated function), but also external borrowing, is championed in Chapter 3.

Facets of diversity are also reflected, albeit in different manners and on a wide variety of topics, in the detailed descriptive studies handled in a typological framework in this volume. Some of the new research findings revealed in this volume are the following:

(i) Based on a corpus of over 300 Sinitic languages, the description and classification of a wide variety of demonstrative systems in Sinitic whose paradigms vary from neutral one-term to highly differentiated types with four or five terms (Y. Chen, Chapter 4).
(ii) The diversity of behaviour with respect to definiteness and word order for bare classifier noun phrases in a sample of over 100 Sinitic languages. Bare classifier phrases are the product of omission of either a demonstrative or the numeral ‘one’ from the position immediately preceding the classifier and its head noun. While this has been described for individual Chinese dialects, it has never been explored on such a scale before (Wang Jian, Chapter 5).

(iii) The highly unusual case in Sinitic of triple sets of pronominal paradigms that are found in certain Northern Wu dialects and include two special series of complex fused forms whose use is determined by specific discourse functions (X.P. Li, Chapter 9).

(iv) The analysis of comparatives of inequality in Southern Min given in Chapter 10 (W. Chen) captures an impressive array of diversity for just this one structure, treated mainly from an intra-linguistic but firmly typological viewpoint for the Min dialects: the Hui’an dialect possesses no less than six different comparative constructions including head-marking, dependent-marking, double-marking, and zero-marking structures. Certain of these can be shown to be the product of innovation or of interdialectal borrowing and also evince diachronic change in progress.

(v) In an enclave formed by a group of different languages and dialects in central and northern Fujian and adjacent areas, an unusual source for the basic numeral ‘one’ is found in the form of GÈ个, which appears to be related to the general classifier in Standard Mandarin. Additional forms for ‘one’ are found as well in this micro-area, all of whose sources derive from words for ‘lone’ or ‘unique’, including a typical Min form, SHŬ蜀. In the Shaowu dialect of Northwestern Min, the numeral GÈ个 operates in complementary distribution with the form, SHŬ蜀 ‘one’, while it also serves as a general classifier, leading to some interesting conjectures on the diachronic development of GÈ个 (S. Ngai, Chapter 8).

(vi) Intra-linguistic variation is considered as an integral part of the description of linguistic diversity for a substantial number of phonological, lexical, and morphosyntactic features in the Southern Pinghua dialects of Guangxi, in addition to language contact issues for the linguistic micro-area they form and share with the Cantonese and Nanning Yue dialects as well as Northern Zhuang (Tai-Kadai/Kra-Dai) languages (H. de Sousa, Chapter 7).

In the chapters which follow, as mentioned, some of the analyses take a macroscopic, typological perspective to analyse large-scale samples of Sinitic languages in terms of their crosslinguistic (or cross-dialectal) variation, while others concentrate on one particular Sinitic language or group of dialects, if not on one particular linguistic area.

For the thematic reasons outlined above, following the two scene-setting chapters in Part I, the large-scale typological studies of Sinitic languages have been united to
form Part II, while the studies on individual languages, language areas, or dialect
groups are presented in Part III.

The unifying thread in this volume, however, regards the findings of all these chapters
which attest to the previously rather meagrely perceived, if not entirely unheeded,
linguistic diversity of Sinitic languages which is slowly but methodically being
uncovered through the work of many scholars, be they in China or in the ‘West’.

The detailed chapter summaries follow:

The first main chapter by Hilary Chappell has as its main objective to show the
great diversity inherent in Sinitic languages from both angles of morphosyntax and
grammaticalization, a diversity which has been little described in the past for Chinese
dialects, apart from the important exceptions of phonology and the lexicon. She
describes the widespread variation in differential object-marking, passive and com-
parative constructions in terms of the lexical source of the marker in the case of the
first two construction types and in terms of both lexical source and structural type for
the comparative of inequality.

Her aim is to see if it is possible to discern linguistic areas within China on this
basis. She tentatively proposes that, in terms of grammatical behaviour, at least five
principal areas exist in China, rather than just a simple North–South division: in
addition to the Northern area, the South is split into the three linguistic areas of
Southwestern, Far Southern, and Southeastern. Furthermore, in between North and
South China lies an intermediate zone, aptly named the Central Transitional area
which demonstrates considerable turbulence in its typological features. Within the
Northern area a small enclave of grammatically conservative Jilu and Jiaoliao dialects
of Shandong peninsula is also to be found. H. Chappell observes that this re
finement of earlier classifications of linguistic areas in China is but the first step, nonetheless,
an undoubtedly necessary first step that can be used as a basis for further research to
verify, re-adjust, or re-align the boundaries.

Alain Peyraube links diachronic grammar and linguistic typology as two import-
ant domains of research on Chinese. He observes that the three basic mechanisms
involved in grammatical change for Sinitic languages are the processes of reanalysis
(including grammaticalization and exaptation), analogy (including the phenomena
of lexicalization and/or degrammaticalization), and external borrowing through
language contact. A. Peyraube points out, however, that at first glance these do not
appear to involve linguistic typology, stricto sensu, or only minimally. The same
applies to the motivations for grammatical change such as semantic-pragmatic
change and structural requirements.

This leads A. Peyraube to consider precisely what connects the two domains.
From the 1980s, diachronic syntax has not ceased to impact on typological
research in China, providing it with a new élan in the form of a framework
upon which hypotheses can be built concerning the common properties which
Sinitic languages share, or the basic differences which separate them.

To illustrate these points, examples are given from Sinitic languages for the
relation between causatives and passives, postverbal and preverbal adverbs, double
object constructions with verbs of giving, and finally, the reanalysis of verbs of saying
into complementizers. All these topics are discussed from a diachronic point of view,
successfully relating this in synchronic terms to typology.

Yujie Chen has undertaken the first large-scale study of demonstrative paradigms
in 303 Sinitic languages, classifying them into neutral one-term, two-term, three-
term, and multiple systems, along the lines of Diessel (1999). She shows that while
two-term systems are widespread in Sinitic, three-term paradigms are not uncom-
mon. In the Gan dialect group whose heartland is in Jiangxi province in Central
China, three-term, four-term, and five-term systems can be found which may use
tone sandhi and reduplication to distinguish the paradigm members for distance:
close-proximal-distal-yonder-further than yonder. Strikingly, one-term neutral sys-
tems may be found to interact with another more complex demonstrative paradigm
in the same language.

In general, these paradigms are sensitive to the parameter of distance from a given
reference point while a small minority are sensitive to visibility as well. Y. Chen also
examines the relation of ‘yonder’ and ‘close’ demonstratives to distal and proximal
categories in terms of implicational hierarchies. She argues that, in the case of Sinitic
languages, ‘yonder’ demonstratives do not necessarily imply the presence of distal
ones, as in certain languages they may be used independently and non-contrastively,
whereas ‘close’ demonstratives consistently imply the presence of a proximal
category.

She concludes by proposing two universals to explain (i) the dominance of the
feature of distance in Sinitic demonstrative systems and (ii) the fact that distal
demonstratives can be differentiated for a greater number of semantic attributes
than are proximal, this being the general case for the other main language families in
Southeast Asia.

Wang Jian presents the results of a typological analysis of bare classifier phrases
based on a sample of 120 Sinitic languages and dialects in order to formalize what has
already been described for isolated cases of Chinese dialects about this phenomenon.
Bare classifier phrases equate to cases where the numeral ‘one’ or a demonstrative has
been omitted, engendering either an indefinite or a definite reading, depending on
certain other factors, which can be explained as follows:

On the basis of the interplay of the two crucial parameters of (i) syntactic position
with respect to the main verb—preverbal or postverbal—and (ii) the interpretation of
definiteness for the bare classifier noun phrase in question, seven main types of
Sinitic languages are identified with respect to the behaviour of this type of noun
phrase. At one end of the spectrum, we find Min dialects which do not allow bare classifier phrases at all in any position, while at the other end, languages mainly located in the Central Transitional zone of China allow all four possibilities for bare classifier phrases. These include Wu, Hui, Gan, Xiang, Hakka, and Jianghuai Mandarin. The unmarked and most common type includes many of the Northern Sinitic languages from both Mandarin and Jin subgroups, but also adjacent languages of Central China. These allow only indefinite, postverbal bare classifier NPs.

Wang Jian concludes his discussion by setting up three important implicational universals connected with these bare classifier phrases for the properties of definiteness versus indefiniteness and pre- or postverbal position. Universal 1 states that if a language possesses preverbal bare classifier noun phrases, it will also possess postverbal ones. Two further universals are related to the tendency in Sinitic languages for preverbal constituents to be definite and postverbal, indefinite, this configuration being the unmarked case: Universal 2 accounts for the fact that if a language allows preverbal bare classifier phrases to have an indefinite interpretation, it also allows for a definite one in this position. Finally, the third universal, Universal 3, accounts for the fact that if a language allows a definite interpretation for postverbal bare classifier phrases, it also allows an indefinite one.

Hilary Chappell and Alain Peyraube set out to explore the interrelationship between synchronic and diachronic aspects of the two main comparative construction types in Sinitic languages, viz, the Type I Compare comparative with \( \text{比} \) and the Type II Surpass comparative with \( \text{过} \) which, broadly speaking, align China on a North–South basis, opposing a dependent-marking strategy to a head-marking strategy. First, they argue for the establishment of a new cognitive schema of a Compare comparative, a schema which has not been attested nor consequently included in any of the main typological studies on this topic, yet represents a crucial kind for Sinitic languages. An overview of the distribution of these two types of comparative structures in China and their markers is next provided with a brief but pertinent digression on the Surpass comparative in the non-Sinitic languages of southern China as well as in those of other Southeast Asian language families.

In the diachronic treatment of this topic, they observe that the earliest written records reveal that Archaic Chinese began as a head-marking Type II comparative language. They show that the Compare comparative is essentially a late Northern development which became dominant in the early Yuan period of the thirteenth century. In the present era, it is steadily encroaching on all major dialect areas within China, and even on the non-Sinitic languages of southern China, competing with the native types, principally the Surpass comparative. They hypothesize that during evolution of the Compare comparative into the major comparative structure of inequality in the north of China, the Surpass comparative was able to independently develop and flourish in central and southern China without any initial competitors.
That it must have had a broader distribution at some earlier time can be seen in the fact that it is still found in peripheral areas, such as in the Jiaoliao and Jilu Mandarin spoken on the Shandong peninsula or in isolated dialects in Shaanxi province, both within the Northern area of China.

**Hilário de Sousa** presents a finely detailed investigation of the interaction between three main languages in the city of Nanning: Nanning Southern Pinghua, Nanning Cantonese—which are both Sinitic languages—and the unrelated Kra-Dai (or Tai-Kadai) language of Northern Zhuang, the major language spoken in the Guangxi Autonomous Region. The prolonged language contact amounts to many centuries for Zhuang and Pinghua speakers, the latter arriving in this area of Southern China as early as the eleventh century. In contrast, there is much less time depth for the Cantonese speakers who arrived relatively recently in the nineteenth century. Nonetheless, the results of this investigation are unexpected in terms of the lexicon versus morphology and syntax. While borrowing freely from Northern Zhuang vocabulary, Pinghua shows much less influence from Zhuang in terms of grammatical features than do the two varieties of Cantonese, Nanning and Standard. Furthermore, Nanning Cantonese proves to be even more Zhuang-like than Standard Cantonese, for which claims of a Tai substratum have been well-substantiated. It is particularly for the ordering of gender affixes with their head nouns, direct and indirect objects in ditransitive (or double object) constructions, and the extent of polysemy in the functions of classifiers, where these differences emerge. De Sousa proceeds to explain these outcomes in terms of a specific set of sociolinguistic parameters conditioning the language contact outcomes.

**Sing Sing Ngai** considers the possible pathways of evolution classifier 个 and its use as the numeral ‘one’ in the Shaowu dialect of Western Min, a dialect spoken in an inland region of Fujian province on the frontier with the Gan dialect group. She treats, in particular, this unusual etymon for ‘one’, which is quite distinct from the corresponding term in most other Sinitic languages, where一 ‘one’ is widespread. A third synonym, constituted by 番蜀 ‘one’ is also discussed. 番蜀 is in fact a common form for the numeral ‘one’ in the Min dialects. All three forms for ‘one’ being found serendipitously to co-exist in the Shaowu dialect, they can be usefully compared for their syntactic distribution, with the added significance of 个 having a classifier usage as well with both nouns and verbs. Several further unusual forms for ‘one’ are discussed for the Central and Western Min region, showing an intriguing diversity for this small area.

Since the frequency and distribution of the numeral use is higher than that of the classifier use, S. Ngai claims that [个] may actually be the indigenous form for ‘one’ in Shaowu which lies in the northwestern corner of Fujian province, while observing that this form, used with the same meaning of ‘one’, may be found in adjacent Wu, Hakka, and Gan-speaking areas.
Ngai argues that the classifier and numeral senses of the morpheme 个 (ki 213) represent a bifurcated pathway of development, and that the numeral use has most likely evolved from an earlier meaning of ‘lone’ or ‘unique’, evidenced in unrelated language families as a common source for the basic numeral ‘one’.

XuPing Li’s contribution concentrates on the Fuyang dialect of Northern Wu to tease apart the argument roles and semantic values of the stressed and the non-stressed uses of complex emphatic personal pronouns. In the Taihu subgroup of Northern Wu dialects, paradigms of complex fused pronouns exist side by side with a simple pronominal series, and include several dialects of the greater Shanghai area as well as Fuyang, Shaoxing, and others. Emphatic pronouns of the complex series in Fuyang Wu are described as having a special discourse prominence, while stress is defined as a strictly acoustically-defined feature.

X.P. Li demonstrates that stressed or emphatic complex pronominal forms are freer in their syntactic distribution than the corresponding unstressed forms. They may occur in a larger number of syntactic positions and construction types, including contrastive topic and focus. Contrastive focus, for example, licenses their use in the basic postverbal object position, but also in the preverbal object position of the KE object-marking construction, as well as in cleft and pseudo-cleft constructions. The unstressed counterparts may not occur in these positions but rather only in preverbal primary and secondary topic positions, although this too effectively crosscuts subject and object argument roles.

X.P. Li also briefly treats diachronic aspects and upholds the view that these complex pronouns are historically the result of fusion of the simple series of pronouns with a preceding copular verb /zi/ in bare cleft constructions. He proposes a cycle of stress placement and focus whereby the originally focus-sensitive complex pronouns are de-stressed and de-focalized, thence undergoing reanalysis to be subsequently used in plain topic positions. In the next stage in the cycle, these complex pronouns are able to receive stress anew, and so may occur once more not only in contrastive topic positions, but also in contrastive focus positions, and thus in a wider range of syntactic positions than that of simply topic—primary or secondary—as is the case for the unstressed emphatic pronouns.

Southern Min languages are well known for their special and often, unique, linguistic characteristics which set them apart from other Sinitic languages. Weirong Chen’s chapter presents no exception in considering the large number of structurally different types of comparative construction found in the Hui’an dialect of Southern Min, on the basis of data gathered from discourse recordings. In all, six different comparatives are analysed in order of frequency of use, being described from the essential angles of syntactic configuration and specific semantic properties but importantly also in terms of typological parameters such as word order harmony and the position of relators. These include both head- and
dependent-marking types, the hybridized or double-marked comparative and the zero-marked type.

The adopted Northern-style 比 Compare comparative proves to be the least frequent in her database while the hybridized structure using both dependent-marking with preposition 比 [pi³] 比 ‘compared to’ and the adverbial head-marking strategy with kha⁵ 恰 is the most common. In fact, three of the Hui’an comparative constructions involve the use of this adverbial marker kha⁵ 恰, which W. Chen hypothesizes may be a case of syntactic reanalysis from a degree adverb ‘a bit, fairly’ to a formal marker of the comparative.

W. Chen is able to reveal some new types that have not been fully described or identified in the typological literature on comparatives and accounts for some of the less frequent structures in the Hui’an dialect in terms of borrowing from Mandarin, and generational change.