In the study of language learning, researchers sometimes ask how languages in contact are related. They compare the linguistic features of the languages, how the mental grammars of each language sub-system are represented, put to use in performance, and how they interact. Within a linguistic family, languages can be closely related or distantly related, an interesting factor, for example, in understanding bilingualism and second language development. Dialects, on the other hand, are considered to be variants of the same language. While there is no way to always draw a sharp line between the categories of language and dialect, it is necessary to distinguish between the two kinds of language variation by the application of uniform criteria. The distinction between dialect and language is important for designing bilingual instructional programs, both for students who already speak two languages and for beginning second language learners.

Keywords: dialect, language contact, language policy, minority languages, biliteracy, Chinese

Introduction

Researchers of language learning do not often delve into questions about how languages and their varieties are related. One topic of debate however, due in part to widespread commentary outside of the field, often presents itself for explanation: the distinction between the categories of language and dialect. Sociolinguists for their part have contributed most notably to the task of distinguishing in a principled way between these categories; and a clear understanding is important for the study of language learning and for the broader discipline of applied linguistics. This article will present the case for continuing to uphold the traditional conception underlying this distinction. Radical qualification and relativistic redefinition only serve to sow confusion among students of language variation and language contact. Clarity is important for research on bilingualism and for second language acquisition (SLA) in particular.
The study of multilingualism in East Asia, China especially, offers the ideal setting for examining the concepts of language and dialect. To begin with, it is important to point out that the general categories of the distinction should apply in the same way as they do in all other situations of bilingualism and linguistic diversity around the world. If it is claimed that they do not, arguments need to be presented why the languages of East Asia are exceptional in this regard.

The occasion that suggested the examination of these issues was an observation made in a review (Xu 2014) of my recent study *Bilingual development and literacy learning: East Asian and international perspectives, BDLL* for short (Francis 2013), regarding the use of the category “language.” The observation was that granting this status to Yue (as spoken in Hong Kong and Guangzhou), Hakka and Minnan (as spoken in Taiwan) and Wu (Shanghai), for instance, does not adhere to objective criteria. Indeed, in the literature on bilingualism, language learning and language contact in China, writers often (although not consistently) mark a dichotomy between national language and regional dialects, with the implied meaning that “dialect” refers to languages that are minority, non-official, lacking standardization and a formal grammar, and an independent and established writing system. In particular this conception is implied when the reference is to a “mere dialect”; for example that in contrast to the national language of China, Putonghua [common speech], Cantonese is a “mere dialect.” In other usages (as in the professional literature from the field of dialectology), all varieties of Chinese, including Mandarin, fall under the category of dialect. Importantly, and in apparent contradiction, languages of the Tibeto-Burman group (of the Sino-Tibetan family) spoken in China are uniformly categorized as languages. In most cases, estimates of communicative efficacy are typically minimized or not taken into account as important in delimiting the categories. As we will see, the inconsistency in official categorization stems from treating branches (of comparable levels) of the Sino-Tibetan family differently, applying criteria that are not uniform from the point of view of objective linguistic criteria, unless, again, it can be shown how and in what way the Chinese branch of Sino-Tibetan is an exception.

If Yue and Hakka are dialects, then we have to say of which language they are dialects. If Yue, Hakka, Minnan, etc. are dialects of the larger category “the Chinese language,” then Mandarin is also one of these dialects, by the same method. But many observers and experts alike would hesitate to call Mandarin, or the official national language based on Mandarin, Putonghua, a dialect. To avoid the confusion, linguists often speak of a group of languages: the “Chinese languages,” the larger branch (in number of speakers) of the Sino-Tibetan family. Sino-Tibetan, in turn, is one of the major phylogenetic units of the world’s languages, a family, among which include the families: Indo-European, Niger-Congo, Afroasiatic, Austronesian, Dravidian, and so forth. So in reverse order (from top-down) one
way of describing the branches of this part of the human language tree is as proposed in Table 1.

Table 1.

<table>
<thead>
<tr>
<th>Sino-Tibetan — family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese — group belonging to the family: Sino-Tibetan</td>
</tr>
<tr>
<td>Yue (for example) — language belonging to the group of Chinese languages</td>
</tr>
<tr>
<td>Mandarin (for example) — language belonging to the group of Chinese languages</td>
</tr>
<tr>
<td>Hakka (for example) — language belong to the group of Chinese languages</td>
</tr>
</tbody>
</table>

Following this schema, dialects in China are varieties, or variants, of these languages.*

* Readers may be familiar with different labels for the categories, for example:
Sino-Tibetan — phylum
The Chinese languages form a family of languages belonging to the phylum: Sino-Tibetan.
Mandarin (for example) — language belonging to the family of Chinese languages
Hakka (for example) — language belong to the family of Chinese languages

Following this schema, dialects in China are varieties, or variants, of these languages.

Mair (1991) adds an intermediary level branch above the level of language in his categorization (dialect → language → branch → group → family) in Table 1. The basic idea is the same.

Every speaker of every language speaks one of its dialects. Thus, a speaker of Mandarin knows one or more of its dialects: Beijing, Zhongyuan, Upper Yangtze, Ji Lu, Northeastern Mandarin, and so forth. If one speaks the variety of Yue from Hong Kong and Guangzhou, we say that he or she speaks the variety, or dialect, of Yue considered to be the most “representative” or “prestigious.” Thus, it is common to make reference to the Cantonese language pointing to or suggesting this representative status (Yue 2015). In such manner, every speaker of a given dialect is a speaker of the language to which the dialect belongs. Or a better way to put it would be: the speaker of a dialect is a speaker of the language to which it belongs by virtue of speaking one of its variants, one of its dialects.¹

Readers familiar with the controversy regarding the categorization of branches and varieties of the Chinese languages will have taken note that the debate often coincides with a political division: writers and researchers working in the Peoples Republic of China (PRC), together with others who favor the official view, on the

¹. The misuse over the years of the term dialect has lead many authors to avoid it (sometimes explicitly) in scholarly reports, a tendency observed in particular among writers who favor the preservation of minority and indigenous languages. Substitute terms include “variety” (Scott and Tiun 2007) or in Latin America, “variante” (INALI 2008), even when it’s obvious that authors are referring to the scientific category: dialect. Labov (2014) is an example of the standard usage in the field (American Standard English and the different non-standard English forms are all dialects).
one hand, and some writers and researchers working outside (e.g., Hong Kong and Taiwan) who present an alternative view. The purpose of this essay is to present the latter, the “orthodox” view, based on established procedures in the field and applied by linguists to other language families, as an alternative framework, even though it clearly represents the minority view among Chinese language scholars today. In other words, my proposal is that the field of Chinese language studies should consider the issues in contention as still open and unresolved.

Years ago, Victor Mair (1991) explained the underlying issues in the diverging conceptions and categories from the perspective of the standard model in language typology. Thus, a good way to begin this reply to the review of BDLL is with a re‑view of Mair’s essay. From one point of view the divergences are terminological, but then they plainly go beyond problems of terminology. In the system of classification of family, group, language and dialect, fangyan [regional speech] is often translated to English as dialect, and conversely, the terms often taken as equivalent in the professional literature. Dialect and fangyan both refer to instances of language variation, and so might coincide in some cases; but the range of meaning for each is not the same. The former is a technical term denoting a relation between two varieties of a language that implies a degree of communicative closeness; that they are not that distant one from the other so as to impede communication. The relationship of relative “distance” between two such dialectal variants of the same language, in essence, is a cognitive (psycholinguistic) estimation: degree of mutual intelligibility. From this point of view, the assessment is about linguistic competence, the mental grammar of the speakers of the respective dialects, or languages, as the case may be. The concept underlying fangyan is formulated differently. In China the speech of two communities may be fangyan, as opposed to separate languages, if they are from the same region, share a common standard language and writing system, and are descendants of a common parent language in history (pp. 5–6). As daughters, a number of interesting linguistic features shared in common can be pointed out that attest to this lineage. A shared cultural inheritance is also implied. As is also the case in other language families, this set of criteria incorporates a wide range of language varieties, including those of speech communities between which there may be only the most approximate and limited mutual intelligibility, if any at all. Mandarin, Yue, Minnan and Wu, for example, meet the standard for fangyan. Dialect, however, implies a different relationship, one of linguistic proximity and intercommunication. On the related point of a shared writing system that transcends dialect and language, the Chinese characters, Mair calls attention to evidence that may undermine this commonly held point of view in its strong version (question to which we will return).
Typology of the languages of China

Returning to the method of classifying the languages that belong to the Sino-Tibetan family in Table 1, a potential imbalance presents itself if we do not apply the traditional method of differentiating between the categories of language and dialect (Mair 1991, 8–11). Sino-Tibetan is divided into two groups: Tibeto-Burman and Chinese. China recognizes a number of national minorities, some of which are speakers of languages belonging to the Tibeto-Burman group (e.g. Tibetan and Yi; Jingpho, spoken in Yunnan province, also belongs to the Tibeto-Burman group). Other minority languages (e.g. Uygur) belong to families outside of the Sino-Tibetan family. In official reports and in the professional literature Tibetan, Yi, Jingpho, Uygur and the other languages of minority peoples are unambiguously categorized as and given the status of separate languages (Li et al. 2014; Rong 2007). This recognition includes the provision of bilingual education and the consideration of related tasks of language preservation (Ma 2007; Sun 2011).

Graphic representations show elaborate branching that reflects the vast diversity of this group. But for the Chinese component it appears that a different standard is applied for levels and branches, implicitly setting the category of Chinese apart, not as a group but as a single language. In contrast, Figure 1 in a highly abbreviated typology, omitting most branches and languages (but the basic idea is the same), applies the same criteria for distinguishing language and dialect for both sides of the family, the same method applied for categorizing languages for rest of the world’s language families. If Tibetan, Yi, Burmese and Jingpho can be shown as separate languages (they are not mutually intelligible), following the same policy-neutral procedure, so can Yue, Mandarin, Wu, Minnan and Hakka.

Figure 1.

The imbalance presents itself in official categorization that implicitly shifts the Chinese category in Table 1 from group to language, relegating the entries
below all to fangyan (translated as dialect). Recall that Putonghua is based on Modern Standard Mandarin, one of the Chinese languages, according to the standard method. For comparison, if we categorize Yue, Mandarin, Minnan, Wu and Hakka as dialects of Chinese, the same approach within the Indo-European family would group Italian, Spanish, Portuguese, Catalan and Romanian as dialects of “Romance”; for discussion, see Cheung and Bauer (2002), Matthews and Yip (2011). Even in the case of the Romance languages, genetically and typologically very close, recent descendants from their parent, mutual intelligibility is nevertheless considerably restricted. For literate speakers, access to a written form in the closely related language often facilities some degree of limited comprehensibility, but only under special circumstances (e.g. predictable expository text on a topic of extensive background knowledge). As an illustration of the above mentioned imbalance, a number of representations of the Sino-Tibetan family that either implicitly or explicitly reject the concept of Chinese-as-language-group (or favor a conception where group=language in the case of Chinese) appear as Figure 2.

![Figure 2.](image)

Policy considerations

What appears evident from the above is that the current official position on the status of minority languages in China flows from a political consideration, one that by the way follows closely that of most other modern nation-states especially

---

2. See the example in Francis (2013, 170–174) of how a limited attainment in reading comprehension can be observed, under special circumstances, in a typologically close language. However, this marginal skill does not imply, in the case cited, that Catalan is a dialect of (a higher order) Spanish (or “Iberian”), political considerations completely aside, on strictly linguistic grounds.
during the historical period of their formation. An assessment of this language policy, the consolidation of a unifying common national language, is the topic in a different forum. It is complex and deserves a separate discussion. But most language planning specialists would probably agree that China will be well served by one aspect of this policy: the universal mastery of Putonghua as the national-official language, soon to be realized among the future generations of school children (Li 2006). The equally complex problem of how and in what way the minority languages of a country can be preserved is also a topic for a separate discussion. Another way to pose this policy question would be to begin with the point of broad agreement: recognition of the benefit of promoting the national-official language, Putonghua/Mandarin, especially in public education, and grant it explicitly and unambiguously the status of language (that is, disfavor the concept of Mandarin as dialect). Achieving the goal of a common language shared by all citizens is neither very controversial nor inherently undemocratic. The alternative view, favored in this reply, would then add the following: realizing this goal does not require that the status of the minority languages of the nation be downgraded, especially if the official categorization on the question at hand turns out not to be consistent. In addition, no intention should be taken from this reply to propose that the term fangyan be discarded. And, if the established consensus regarding the categories used in Chinese dialectology is important for overriding political constraints, such a position does not make international scholarly exchange impossible, far from it in fact. Each point of view (competing schema, so to speak) can simply take into account the difference in terminology and conception of the relevant categories.

Then there are the political considerations of other speech communities. Two in particular have come forward in recent years: on the part of advocates for pluralism and recognition of linguistic diversity in Taiwan, and for greater democracy and autonomy in Hong Kong. Speakers of Minnan and Hakka in the former and Cantonese in the latter (the majority languages in each case) have increasingly came to view their languages as such, not as simple dialectical variants of an overarching Chinese language embodied and officialized in Putonghua (Guoyu in Taiwan). With the lifting of martial law in Taiwan in 1987, came the eventual abrogation of enforced restrictions associated with the imposition of Modern Standard Mandarin as the sole language of schooling, in all domains, including informal uses among students and teachers. With time, the Mother Tongue Language Movement, despite its limitations (Liu 2012), has reversed the perception of Minnan, Hakka and the Taiwanese indigenous languages as unworthy of official recognition (Chen 2010; Wu 2012). Today they are recognized by the majority of the population, together with Mandarin, as languages of the nation, despite
severe obstacles that continue to block normalization. In a similar manner, the educational language policy in Hong Kong is **biliterate** (Chinese and English) and **trilingual** (Putonghua, English and Cantonese), not bilingual+one dialect. Young people in Hong Kong today, active in the Umbrella Movement, would bristle at the suggestion that in regard to their bilingual competence in Chinese that one linguistic entity is a dialect and the other is a language; and that the first should cede its place to the second. While Tibetan and Uyghur are granted language status (as languages of national minorities), analogous considerations enter the discussion regarding their recognition and use within Tibet and Xinjiang — for example, the implementation of effective models of developmental bilingual education (Wan and Zhang 2007). Professional linguists are now placed in a difficult position: in the face of an informed bilingual colleague, speaker of a minority language, should they accept the validity of his or her scientific argument in one context and deny that the same evidence is pertinent in another?

3. An anonymous reviewer correctly called attention to the difficult, even precarious, revitalization effort on behalf of Minnan in Taiwan. On the one hand, the failure to date to reach consensus on an orthographic representation weighs heavily on the continued lack of progress in normalizing its use. This factor is only one among a broader array of adverse conditions (only a small percentage of young Taiwanese today do not speak Mandarin, the other autochthonous languages have received official recognition since democratization and vie for preservation); see and Chang and Lu (2014), Dupré (2014) and Liu (2012). Nevertheless, such language status deficits and sociolinguistic imbalances that strongly tip the balance in favor of Mandarin have no effect, one way or the other, on the proposal to categorize Minnan as a language, even in the hypothetical case of future advanced erosion.

4. By the same token, linguists are under no obligation to endorse the artificial division and exaggeration of differences among mutually intelligible dialects (e.g. in Europe) for the purpose of declaring them as separate languages. Again, scientific criteria override politically motivated categorization. Of course, there are numerous examples of truly intermediary grouping along psychologically real continua, topic for another day.

Why is this discussion important for the field of bilingualism and second language acquisition?

Not all school children and adult students in China are L1 speakers of the national language, similar to the situation of linguistic diversity of most other countries in the world. Thus, as is no different elsewhere, educators and investigators need clarity on basic concepts of language variation for the purpose of properly framing the research problems related to academic achievement and second language proficiency for L2 learners of Putonghua/Mandarin. In particular, the problem of
literacy learning in a language that is not the learner’s L1 requires this clarity. More broadly, bilingualism in school and in society is widespread. The important research questions posed by this knowledge of more than one language are the same as those in other Asian countries, and then beyond. A clear definition, to include all minority languages (not only the officially recognized languages of national ethnic minorities) will help make the evaluation of findings from studies in SLA more consistent and coherent. Is the student who speaks Mandarin and Tibetan “bilingual” while the one who speaks Mandarin and Cantonese “bidialectical” for a linguistic science reason or for a policy reason?

Returning to Mair’s (1991) essay, two important questions are posed for the field of applied linguistics in China centered specifically on problems of bilingualism and L2 learning: Modern Standard Mandarin as L2 and L2 literacy learning of Modern Standard Written Chinese (MSWC). Of course, the same applies for L1 speakers of Mandarin who need to learn a non-official regional language.

(1) What exactly is the nature of the grammatical differences, regarding both morphology and syntax, among the different Chinese languages (fangyan, if the reader prefers)? The widely accepted claim that they differ mainly in vocabulary, pronunciation and idiomatic expression with a shared grammar still needs to be systematically evaluated (p. 7). Such a comparative linguistic analysis could be carried out in parallel to the related psycholinguistic study: the controlled assessment of mutual intelligibility between monolingual subjects. Tang and van Heuven (2015) propose a procedure for estimating relative distance.

(2) The claim that in the Chinese writing system character-recognition freely crosses linguistic boundaries requires further empirical scrutiny. Anecdotal evidence and observation suggest the need to test the assertion, case by case, that speakers of all languages and varieties of Chinese are able to learn in the same way the morphosyllabic characters independent of their linguistic competence (Mair 1991, 14). The relevant test of this claim involves the reading ability of students who are not speakers or intermediate level learners of Putonghua/Mandarin. In other words, the reading ability of bilinguals or intermediate level learners (e.g. L1 Cantonese/intermediate L2 Mandarin) would not count as relevant evidence for the Chinese-character-as-written-lingua-franca hypothesis. Observational evidence, reported by Mair (2002), from the experience of developing vernacular writing systems (e.g., for Taiwanese and Cantonese) tentatively suggests that the strong version of the written-lingua-franca hypothesis is not correct.

Studies of the spontaneous development of literacy in Cantonese and the evolution of a vernacular non-standard writing system point to additional indirect evidence
that literacy in Chinese characters is not language-independent. Far from the popular conception of a language-neutral, “logographic” script with direct-access-to-meaning, the processing of characters in reading activates all the components of linguistic competence, including phonology, as is the case of reading in alphabetic writing systems (He et al. 2005; Perfetti and Liu 2005). It appears from anecdotal report that skilled reading of standard Chinese characters is facilitated by a certain level of linguistic competence in Putonghua/Mandarin, and similarly, for the vernacular Cantonese script. In their study of written Cantonese, Cheung and Bauer (2002, iii) argue that: “[The] Cantonese language is not the standard Chinese lexicon cloaked in Cantonese phonology…The Mandarin speaker who does not know Cantonese may find a text of Cantonese almost unintelligible.” Nota bene: no suggestion should be taken from the above that all versions of the written-lingua-franca hypothesis are incorrect. The Chinese writing system is unique among the world’s orthographies, and the research on the processing of text written in Chinese characters is still in its infancy. As such, the question of their capacity to “cross” linguistic boundaries (in contrast to this capacity, lack of rather, among alphabetic systems) needs to be considered case by case. In particular, the unique morphosyllabic feature of characters poses this difficult and very interesting question. For example, to what extent, and if how, are they capable of this “crossing” in the case of varieties or languages close to Putonghua/Mandarin and varieties or languages distant? Li (2006) addresses this research problem. It’s not a simple problem.

A number of interesting research reports have highlighted the learning and teaching challenges for L2 school-age populations (Poon 2010; Yeung et al 2013). In Hong Kong, for example, despite the introduction of Putonghua into the curriculum as core subject since 1998 (Leung 2005), progress in actual student proficiency appears slower than expected. Difficulties are encountered that are typically associated with the learning of a “foreign language” (Li 2009, 81). An argument presented for improving Putonghua instruction is that it is more consistent with the standard Chinese characters (MSWC), thus facilitating literacy learning for Cantonese-speaking students. The argument, interestingly, is made from two points of view (Tam 2011, and from a different perspective, Tse et al. 2007).

The paper by Tse et al. (2007) is noteworthy because it predicted strong negative transfer, “interference” (p.403), from students’ L1, Cantonese, when learning to read in the national language. The great majority were L2 learners of Putonghua. The conceptual framework of the study is of note because the authors lean toward a deficit-model approach in the introduction and literature review. Considered as a “mere dialect” they might perhaps object to categorizing the influence of Cantonese as a “L1” factor. The favored model strongly put forward by the authors is the replacement of “Cantonese with Putonghua as the medium
of instruction when teaching Chinese, that is if teachers are linguistically competent and students linguistically able” (p. 403). Pointing to reported difficulties in learning Putonghua and in mastering MSWC, one possible causal factor proposed was that students “habitually speak Cantonese” and are “denied opportunities to become familiar at first-hand” with the national language upon which MSWC is based (p. 403). A focus of the large-scale study (N=4,335) was the immigrant population from the Mainland previously educated in Putonghua, with literacy learning directly in MSWC. The assumption was that their vocabulary and grammar knowledge would not have been “contaminated by Cantonese and their linguistic skill… superior to that of their peers and even, in some cases, to that of their teacher” (p. 405). The researchers sought to compare (primary class 4) children with different amounts of exposure to Putonghua at home within a population where approximately 90% spoke Cantonese as the home language. The main performance index, appropriately, was a measure of reading comprehension. Interestingly, test results did not confirm the (informal) hypothesis of Cantonese contamination; in fact, the opposite result was obtained. Even with superior scores for Mainland-born students, “…those frequently using Cantonese at home displayed significantly better performance (p<0.001) than counterparts only speaking Putonghua at home. In fact, against the expected outcome, students who always speak Putonghua at home achieved the lowest average score… Students who ‘sometimes’ speak Putonghua at home performed the best on average…” (p. 407). “Those who always use Cantonese at home language gained the highest reading scores” (p. 409). A full review of the study’s research design and assessment procedures needs to be deferred to another occasion. For now, we can conclude from the findings that the authors’ implicit dismissal of Hong Kong’s biliterate and trilingual education model — Poon (2010) and Wang and Kirkpatrick (2013) provide a comprehensive overview — should be reconsidered. Internationally, correctly implemented bilingual literacy instruction has been shown to be highly productive and efficient, with no disadvantage in ultimate attainment.

The growing consensus in research on bilingual development, surprising as it was at first, is that each of the two language subsystems achieves a neurocognitive separation from an early age, and that this autonomy is preserved (Francis 2012; Matthews and Yip 2014). Nevertheless, the language subsystems maintain an extensive network of interfaces, allowing for all kinds of interaction and

5. Importantly, students’ actual knowledge of Putonghua was not independently evaluated in the Tse et al. study. Thus, the findings do not bear on the question of linguistic competence (level of L2 knowledge) and L2 literacy learning. But indirectly, the results are not consistent with the claim of a significant interference effect that Cantonese speakers might be burdened with in school.
mutual transfer: as in cross-language influence ("transfer") on many levels, and in codeswitching. This finding (still open to discussion) is an example of how the assessment of results from studies of student achievement in language-related domains will benefit from a more complete integration into the global exchange among investigators. This assessment will be more clearly framed by concepts from the fields of second language acquisition and bilingualism internationally. Instruction that brings together Cantonese and Putonghua, or Taiwanese and Guoyu, could be designed more deliberately as bilingual education that implements the best practices of dual language and second language teaching.

In the short term we can agree to disagree on questions of terminology and concept and simply acknowledge that for many researchers dialect should not be taken as the correct translation of fangyan. This approach is in line with the compromise solution proposed by Dong (2014, 155–156). Taking the different perspectives and constraints into account, specialists who work on these problems can continue the discussion on language variation without having to try to disqualify the opposing view on this one point. But as an anonymous reviewer suggested, postponing the debate forever will not be possible. Further research involving measurable indices of mutual intelligibility and the findings from a broad sample of the Chinese branch of the Sino-Tibetan family needs to continue. This research will eventually be held up in comparison to how the terms “language” and “dialect” are understood in the field as a whole. The goal will be to arrive at a unified conception.

References


**Author’s address**

Norbert Francis
Northern Arizona University
Flagstaff, Arizona
USA

Norbert.Francis@nau.edu
Copyright of Chinese Language & Discourse is the property of John Benjamins Publishing Co. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.