

## **Addenda and Corrigenda**

- the talk presenting the “arbre” by Laurent Sagart can be accessed online at: URL: <https://www.academia.edu/19534510/>
- I apologize for the bad quality of the graphics (especially on the Brill website, since this author's version is actually better), which is very surprising for me, since I submitted all graphics in vector format and in high-resolution JPG, but somehow the production team was incapable of realizing that one cannot downscale graphics at libitum. I learned my lesson for the future and will from now on no longer accept any proofs until I have the complete confirmation that the graphics are rendered exactly as I submit them.

# Network Perspectives on Chinese Dialect History

## *Chances and Challenges*

*Johann-Mattis List*

CNRS-CRLAO-INALCO-EHESS

*mattis.list@lingpy.org*

### Abstract

Little is known about the history of Chinese dialects. Major dialect groups were identified long ago using various traditional criteria, such as tonal and segmental development from their presumed common ancestor; however, scholarly agreement about their detailed development is largely lacking. At the core of the problem lies the role that language contact played in the history of Chinese. Unlike in the case of other language families, the Chinese dialects never really separated into distinct, independent languages, but kept evolving in close contact to each other. As a result, it is hard to tell whether traits shared among the dialects have been inherited or borrowed. Recent network approaches from a biological perspective could show a way out of this dilemma, since they were specifically designed to handle vertical and horizontal aspects in bacterial evolution, and the first pilot studies in historical linguistics have reported promising results. In this paper, a case study on the application of network approaches in Chinese historical linguistics is presented. Based on a dataset of 200 basic items translated into 23 Chinese dialect varieties, competing proposals for Chinese dialect classification are compared and tested for general plausibility. The results of the comparison show that network approaches are a useful supplement for quantitative and qualitative approaches in Chinese historical linguistics. In order to reach their full potential, however, the underlying evolutionary models need to be more closely adapted to linguistic needs, and additional evidence, like geographic information, needs to be taken into account. (This article is in English.)

### Keywords

Chinese dialect classification – Chinese dialect history – network approaches – lexical borrowing

## 1 Chinese Dialect Classification and Chinese Dialect History

The sociolinguistic situation in China is incredibly complex: For a long time, countless, mostly mutually unintelligible, varieties of Chinese have been developing under the backdrop of a common culture and writing system. Influenced by centripetal forces of varying prestige languages, centrifugal forces of geographic distance, and different waves of migrations, the modern Chinese dialects constitute a language family whose linguistic divergence resembles that of the Romance languages (Norman 1988: 187–188, Wang 1997), but whose history is so complex and intertwined that it seems impossible to describe it by means of a classical family tree (Norman 2003: 76).

### 1.1 *Traditional Approaches to Chinese Dialect Classification*

Figure 1 shows the geographic distribution of the ten major dialect groups proposed by Wurm and Liu (1987).<sup>1</sup> Despite a long tradition of research, there is little agreement among scholars regarding the exact classification of the Chinese dialects. Although the majority of scholars probably agrees with the distinction of the seven major groups of Mandarin (Guānhuà 官話), Xiāng 湘, Gàn 贛, Wú 吳, Hakka (Kèjiā 客家), Yuè 粵, and Mǐn 閩 (Norman 1988: 181), there is less agreement regarding the status of the three more recently proposed groups of Jin 晉, Huīzhōu 徽州, or Pínghuà 平話 (Kurpaska 2010: 64–73, Yan 2006: 8–19), and there is even less agreement regarding the further subgrouping of the major groups.

Traditionally, Chinese dialect classification is based on the comparison of the sound system of Middle Chinese with the sound systems of the modern dialects (Yan 2006: 9). Assuming that all Chinese dialects more or less go back to the language variety encoded in the extant rhyme books published in the 6<sup>th</sup> century A. D. (traditionally labeled 'Middle Chinese' or *Zhōnggǔ Hànyǔ* 中古漢語), it is straightforward to determine how specific features, such as the voiced plosives, or tonal categories, are reflected in a given modern dialect variety. In order to compare the features of a given dialect variety with those of Middle Chinese, scholars usually rely on character readings. Based on the pronunciation of diagnostic characters in the target dialect, the phonological differences, compared to the Middle Chinese system, are determined and employed for classification.<sup>2</sup> Table 1 gives examples for dialect readings of characters with Middle Chinese labial stops as the initial (*b, p, p<sup>h</sup>*), with each of the dialect varieties representing one of the seven major dialect groups (data taken from Běijīng Dàxué 1962 [1989]).



FIGURE 1 Geographic distribution of the ten major dialect groups of Chinese.

1 Figure with modifications taken from [http://en.wikipedia.org/wiki/File:Map\\_of\\_sinitic\\_dialect\\_-\\_English\\_version.svg](http://en.wikipedia.org/wiki/File:Map_of_sinitic_dialect_-_English_version.svg).

2 For a typical diagnostic character list, compare the *Fāngyán diàochá zìbiǎo* 方言調查字表 (Zhōngguó Shèhuì Kēxuéyuàn 1955[1981]).

TABLE 1 Comparing character readings in the major Chinese dialect groups

Character	Middle Chinese	Běijīng 北京 (Mandarin)	Sūzhōu 蘇州 (Wú)	Chángshā 長沙 (Xiāng)	Nánchāng 南昌 (Gàn)	Měixiàn 梅縣 (Hakka)	Guǎngzhōu 廣州 (Yuè)	Fúzhōu 福州 (Mín)
pó 婆 'old woman'	<b>b</b> a	p <sup>h</sup> o <sup>35</sup>	<b>b</b> u <sup>24</sup>	p <sup>h</sup> o <sup>13</sup>	p <sup>h</sup> o <sup>24</sup>	p <sup>h</sup> ɔ <sup>11</sup>	p <sup>h</sup> o <sup>21</sup>	p <sup>h</sup> ɔ <sup>52</sup>
bō 波 'wave'	<b>p</b> a	p <sup>h</sup> o <sup>55</sup>	<b>p</b> u <sup>44</sup>	p <sup>h</sup> o <sup>33</sup>	p <sup>h</sup> o <sup>42</sup>	p <sup>h</sup> ɔ <sup>44</sup>	p <sup>h</sup> ɔ <sup>53</sup>	p <sup>h</sup> ɔ <sup>44</sup>
pō 坡 'slope'	p <sup>h</sup> a	p <sup>h</sup> o <sup>55</sup>	p <sup>h</sup> u <sup>52</sup>	p <sup>h</sup> o <sup>33</sup>	p <sup>h</sup> o <sup>42</sup>	p <sup>h</sup> ɔ <sup>44</sup>	p <sup>h</sup> ɔ <sup>53</sup>	p <sup>h</sup> ɔ <sup>44</sup>

### 1.2 Current Challenges in Chinese Dialectology

In the last twenty years, more and more scholars have begun to question the traditional approach to Chinese dialect classification, demanding a paradigm shift in Chinese dialectology (Norman and Coblin 1995, Branner 2000, Baxter 2006). The traditional classification criteria are at the core of the problems. The tendency to reduce rather than to increase the number of features used in dialect classification is “essentially arbitrary” (Baxter 2006: 75), since the classification will directly depend on the initial choice of features (*ibid.*). Feature reduction will also necessarily lead to a loss of information and deprive genetic classifications of their discriminating power. The reflexes of Middle Chinese initials may be sufficient to identify most standard dialect groups (Lǐ 2005). However, in such a classification, the features are merely used to *distinguish* certain dialect groups, while they do not *explain* how they developed.

Although most classifications proposed so far are explicitly based on historical criteria, few of them explicitly try to account for the genealogical development of the Chinese dialects (Yóu 1992: 85–114, Norman 1988: 210–214, Wáng 2009, Sagart 2011). If the goal of classification was only a partitioning of the dialects into different groups that bear no historical implications, it would be much more useful to employ synchronic criteria, such as mutual intelligibility (Yóu 1992: 49–50, Zav’jalova 1996: 18), to accomplish this goal. However, if the purpose of dialect classification is to infer ‘relationships of common ancestry’ (Baxter 2006: 75), it is obvious that the classification needs to be based on features which are diagnostic for genetic relations. Here, most scholars opt for the use of *lexical features*—namely *basic*, or *core vocabulary* (Norman 1988: 181–183, Branner 2000: 25–28, Sagart 2011), since the borrowing of core vocabulary ranks rather low in the standard ‘scale of borrowability’ (Aikhenvald 2007: 7). However, since vocabulary can generally be easily exchanged (Weinreich 1953), basic vocabulary is no miracle cure against the “horizontal forces” of language history. In order to improve Chinese dialect classification, it seems inevitable to improve on those methods that can help to distinguish *borrowed* from *inherited* traits.

### 1.3 Phylogenetic Networks

Given the complex sociolinguistic situation in China, it is not surprising that many scholars claim that the *family tree model* (first proposed by Schleicher 1853, see Figure 2A) is inadequate to model Chinese dialect history, since it ignores the horizontal (areal) dimension of language relations that has played such an important role for the development of the dialects into their current shape (Sagart 2001, Norman 2003). Unfortunately, the alternative model, the *Wave theory* (Schmidt 1872, see Figure 2B), is also not very helpful, since it ignores the vertical dimension of language relations that is—of course—also constitutive for the history of Chinese dialects. Network models show a way out of the dilemma, since they can be easily used to display both *vertical* and *horizontal* language relations (see Figure 2C), as illustrated early on by Southworth (1964) for the Indo-European languages, and in a recent paper by Wáng (2009) for the Chinese dialects. In the following section, I will try to illustrate how methods for phylogenetic network reconstruction, which have their origin in evolutionary biology but were recently adapted to linguistic needs, can be used to test and explore hypotheses on Chinese dialect classification.

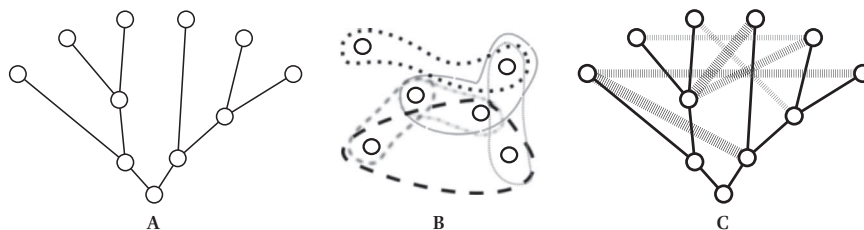


FIGURE 2 Trees (A), waves (B), and networks (C).

## 2 Reconstruction of Phylogenetic Networks

### 2.1 Distance- and Character-Based Approaches

It is common to make a distinction between *distance-* and *character-based* methods for phylogenetic reconstruction. The main difference between these different families of methods lies in the aggregation of information: distance-based methods aggregate information on the taxonomic level. Similarities and differences between all taxonomic units (language varieties) are reduced to distance scores. Character-based methods aggregate information on the level of the items that are selected to define the taxonomic units. Character-based methods yield concrete, individual evolutionary scenarios for each character in the dataset.<sup>3</sup>

The most popular distance-based methods for phylogenetic network reconstruction are based on techniques that produce *splits networks* (Huson et al. 2010: 87–126). The most popular algorithms are *split decomposition* (Bandelt and Dress 1992) and *NeighborNet* (Bryant and Moulton 2004). For both algorithms, the popular *SplitsTree* software package offers user-friendly implementations (Huson 1998). Splits networks are quite popular in historical linguistics and have been used in many studies on different language families (McMahon et al. 2005, Ben Hamed 2005, Ben Hamed and Wang 2006). However, the new insights these methods provide are rather limited. Only very general conclusions regarding the tree-likeness of the data can be drawn and the results are extremely difficult to interpret. Rates of borrowing cannot be calculated, nor can individual borrowing events be inferred.

The limited information provided by splits networks is illustrated in Figure 3, where a splits network of ten Chinese dialect varieties is displayed.<sup>4</sup> As can be seen from the network, there seems to be a considerable amount of conflicting signals. Wēnzhōu 温州, a Wú dialect, for example, appears in an intermediate position between the Wú (Shànghǎi 上海) and the Mǐn dialects (Fúzhōu 福州 and Táiběi 台北). Chéngdū 成都, a South-Western Mandarin dialect, is placed between Shànghǎi and Chángshā 長沙 (Xiāng) dialects. However, the network does not show us *why* some varieties occur in conflicting positions, and it also does not show the *historical processes* underlying the conflicts.

Character-based methods for phylogenetic network reconstruction cope with many shortcomings of distance-based approaches. Since they handle the evolution of individual characters (like cognate sets, or phonetic features), they offer distinct hypotheses regarding character development and offer a unified representation of the tree- and wave-like components of language history. These methods are

3 Note that the conversion to distances or characters is crucial for all approaches making use of bioinformatic software. The nature of the original data, be it words, cognate sets, or even structural data (as used, for example, in Dunn et al. 2005).

4 The network was reconstructed with help of the NeighborNet algorithm applied to a small excerpt of the dataset by Hóu (2004) as prepared for the study by List et al. (2014). The network was visualized with help of the SplitsTree software package (Huson 1998). The original dataset and the distance matrix fed to the algorithm is available in Supplementary Material 1.

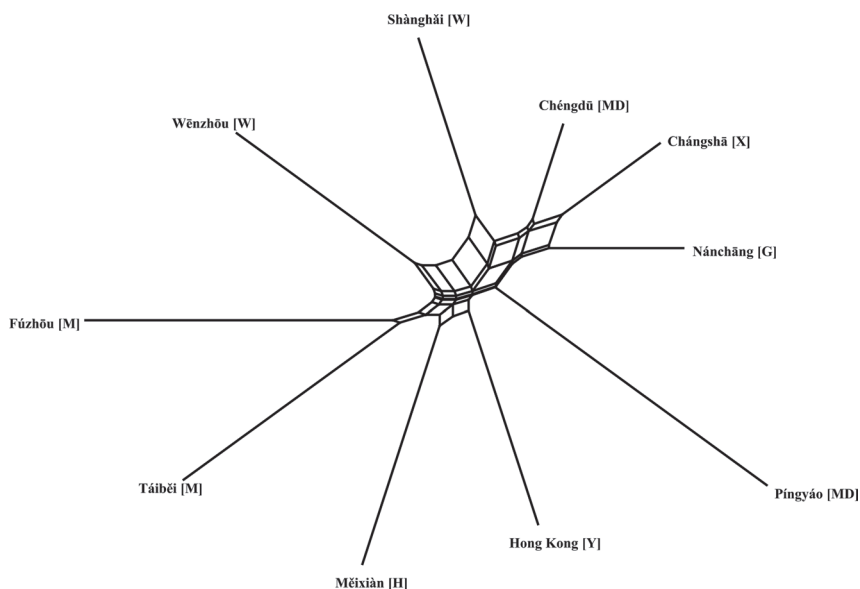


FIGURE 3 Splits network of 10 Chinese dialect varieties (see text for details regarding the data-set and the analysis).

very common in evolutionary biology (Huson and Scornavacca 2011, Koonin et al. 2001). In linguistics, however, there are only a few proposals so far (Wang and Minett 2005, Nakhleh et al. 2005). Unfortunately, none of these approaches are publicly available, be it in form of standalone software tools or as software libraries. This makes their replication and application to new datasets extremely difficult and tedious. An alternative character-based method, which is freely available, is the so-called *minimal lateral network* (MLN) approach (Dagan and Martin 2007, Dagan et al. 2008, Nelson-Sathi et al. 2011, List et al. 2014 a and b), which is implemented as part of a Python library for quantitative tasks in historical linguistics (List and Moran 2013, List et al. 2015). The MLN approach is a character-based method for phylogenetic network reconstruction. The method was originally designed to study microbial evolution. The basic idea of the approach is to test how well a given phylogeny (a so-called *reference tree*) explains the evolution of a set of characters. Instead of simply testing the explanative force of phylogenies, the MLN approach tests different models in which increasing amounts of horizontal transfer are allowed, and then it selects the model that provides the best explanation for the evolution of the characters. In a pilot study by Nelson-Sathi et al. (2011), the MLN approach was used to assess borrowing frequencies during Indo-European language history. In List et al. (2014a), an improved version of this approach was presented and successfully applied to a small set of 40 Indo-European languages, where it identified 72% of all known borrowings in the data. In List et al. (2014b), the method was applied to Chinese dialect data, where it revealed a much higher amount of characters suggestive of borrowing than was inferred for Indo-European (48–55% in Chinese vs. 31% in Indo-European).

The benefits of character-based network approaches as opposed to distance-based approaches are illustrated in Figure 4, where the same ten dialect varieties were analyzed with the help of the MLN approach.<sup>5</sup> As can be seen from the Figure, the network yields a direct hypothesis regarding the

5 The minimal lateral network was reconstructed using the same dataset as used for the NeighborNet analysis and the most recent implementation of the MLN method provided as part of the LingPy software package (List and Moran 2013, Version 2.4.1-alpha, List et al. 2015). The original dataset and a script that reproduces the analysis is available in Supplementary Material 1.

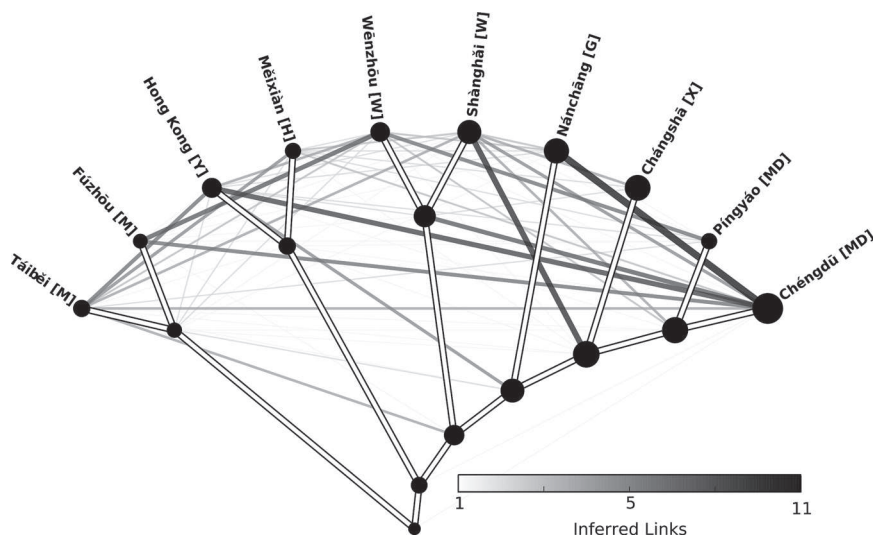


FIGURE 4 Minimal lateral network of 10 Chinese dialect varieties (see text regarding details of dataset and analysis).

intermediate position of Wēnzhōu and draws a horizontal edge between Fúzhōu and Wēnzhōu, thus indicating that these two varieties share a considerable amount of characters which they do not share with other varieties. It also shows that the position of Chéngdū in the splits network can be more readily explained as having a specific contact relation to Nánchāng 南昌, and a more ancient contact relation between the Shànghǎi dialect and the ancestor of Chéngdū, Píngyáo 平遙, and Chángshā dialects. We note further that the network has a *root*, and *internal nodes*, representing the ancestors of the contemporary varieties. The network thus contains an explicit *temporal* dimension. The varying size of the nodes further indicates how many characters (word forms for this specific dataset) are attested for each of the contemporary varieties, and how many word forms are inferred for ancestral varieties.

## 2.2 Minimal Lateral Networks

The MLN approach (List et al. 2014 a and b) takes as input a *reference tree* and a set of *phyletic patterns*. Phylogenetic networks are inferred within a three-stage approach. In the first stage (1), *gain-loss mapping* techniques are used to infer a range of different *gain-loss models* that explain how the characters could have evolved along the reference tree. In a second stage (2), the best model is chosen by comparing the ancestral and contemporary *vocabulary size distributions*. In the third stage (3), a *minimal lateral network* is reconstructed from the gain-loss scenarios inferred by the best model.

*Gain-loss mapping* (GLM, Cohen et al. 2010, Mirkin et al. 2003, List et al. 2014 a and b) is a standard technique in evolutionary biology. It can be used for various purposes, such as the assessment of the tree-likeness of a given dataset, the inference of horizontal gene transfer events, or the reconstruction of ancestral character states. The basic goal of all GLM approaches is to infer *gain-loss scenarios* that explain how a given *phyletic pattern* developed along a reference tree. A *phyletic pattern* is a matrix representation of the distribution of cognate sets in a given set of languages. The matrix displays whether cognate sets have *reflexes* in a given language or not. For each language, a given cognate set is represented by two states: presence (1) or absence (0). Depending on the cognate sets being investigated, different patterns can be observed. This is illustrated in Table 2 where translations of ‘to count’ in three Romance



and three Germanic languages are split into two cognate sets and coded as phyletic patterns. Given a reference tree that reflects the general evolution of languages, a gain-loss scenario (GLS) explains the evolution of a character in terms of *events*. Events are state changes from ancestral to descendant nodes of the reference tree, with gain events being defined as changes from state 0 to state 1, and loss events being defined as changes from state 1 to state 0. In lexicostatistical terms, a loss means that a given word form is no longer used to express a given concept. It does *not* mean that the word is completely lost from the language. As in the case of ‘count,’ a possible English cognate for the German and the Danish form would be *tell*, which goes back to the same Proto-Germanic root. However, since *tell* ceased to denote the action of ‘counting things,’ it is not included in our sample, and its current state in English is represented as absent (0) in the phyletic pattern in Table 2.

Figure 5 shows two possible gain-loss scenarios for the first phyletic pattern from Table 2. In scenario A, one gain event and two loss events are inferred. The scenario thus implies that English *count* was inherited from the common ancestor of the Romance and the Germanic languages with its reflexes being lost in German and Danish. Scenario B, however, implies that no ancestral form of *count* was present in the common ancestor of the Germanic languages and that it originated independently in English and the ancestor of the Romance languages. We know, of course, that the second scenario is the right one, since English *count* was borrowed from Old French *conter*. Given that the independent origin of characters in different branches of a language family is rather rare, we can make the (simplifying) assumption that *patchy cognate sets*, i.e. cognate sets for which a given gain-loss scenario suggests *multiple gain events*, are the result of language contact.

TABLE 2 Phyletic patterns for ‘to count’ in Romance and Germanic languages

	Spanish	French	Italian	English	German	Danish
‘to count’	<i>contar</i>	<i>compter</i>	<i>contare</i>	<i>count</i>	<i>zählen</i>	<i>tælle</i>
Latin <i>computare</i>	1	1	1	1	0	0
Proto-Germanic <i>*tal-</i>	0	0	0	0	1	1

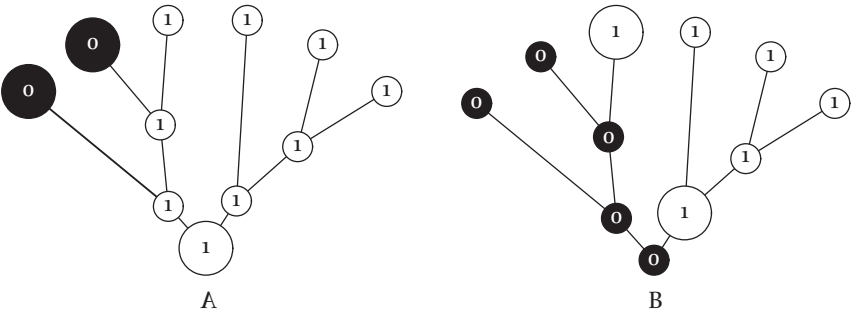


FIGURE 5 Gain-loss mapping analyses of reflexes of Latin *computare*. A shows analysis for one gain and two losses. B shows the analysis for two gains and no inferred loss events (see text for further clarification).



In order to find an appropriate gain-loss scenario for a given phyletic pattern, it is important to find criteria that define the appropriateness of gain-loss scenarios. To know whether the second of the two scenarios in Figure 5 is the right one does not help us to select the correct scenario in cases where we do not know the history of the languages in such great detail. Internally, it is easy to define gain-loss models that favor one of the scenarios by either restricting the maximum number of gain events (*restriction-based approaches*, Nelson-Sathi et al. 2011), or by defining specific “penalties” for gain and loss events (*parsimony-based approaches*, List et al. 2014 a and b). Externally, however, specific criteria are needed to determine the best model for a given dataset. Nelson-Sathi et al. (2011) follow Dagan and Martin (2007) in using ancestral vocabulary size distributions as a heuristic to determine an optimal gain-loss model. The vocabulary size distribution (VSD) of a given language is defined as the number of words the language uses to express a given set of concepts. The basic idea of this approach is that the number of words that are used to express a given number of concepts in ancestral languages should not differ greatly from the number of words used to express the same concepts in contemporary ones. As illustrated in Figure 6A-C, models that overestimate the tree-likeness of the data yield ancestral VSDs that grow drastically (A), while models that overestimate the amount of horizontal transfer yield drastically shrinking VSDs (B). The preference should be given to models that yield well-balanced VSDs throughout all nodes of the tree (C).

Once an appropriate gain-loss model for a given dataset has been determined, the results of the GLM analysis can be interpreted and further analyzed in different ways. The simplest way is to sort out all patchy cognate sets and to investigate these cases individually. Patchy cognate sets can have different origins: They may result from (a) independent convergent evolution, (b) any form of contact between the involved languages or their descendants, including direct, but also semantic transfer, or (c) errors in the data. Given that independent convergent evolution is not a very frequent process (neither in biology nor in linguistics) and that errors in the data should not occur in an ideal world, it is straightforward to assume that the patchiness of the cognate sets results from contact. For a global representation of all patchy cognate sets inferred for a given dataset, one can reconstruct a rooted phylogenetic network (Dagan et al. 2008, Nelson-Sathi et al. 2011) which displays patterns of vertical and horizontal inheritance. The reference tree represents vertical relations and serves as the basic hypothesis regarding the genetic classification of the language varieties under investigation. Additional edges drawn between the nodes of the tree represent the number of times multiple gain events were inferred and indicate potential contact relations between the respective nodes or their descendants in the tree.

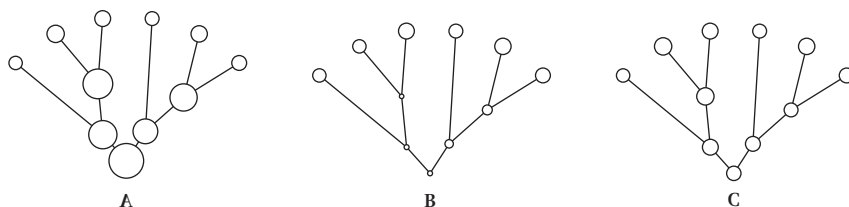


FIGURE 6 Vocabulary size distributions for different gain-loss models. A shows a growing VSD for models which do not allow for horizontal transfer. B shows a shrinking VSD for models that overestimate the amount of horizontal transfer in the data. C shows a moderate VSD for models that allow just for the right amount of both vertical and horizontal processes in language history.

### 3 Minimal Lateral Networks of Chinese Dialects: A Case Study

Few people will deny that neither *Stammbaum* nor *Welle* are sufficient to reflect Chinese dialect history in all its complexity. But do network approaches at their current stage offer a serious alternative to the two well-established models of language classification? In the following section, the application of network approaches in Chinese dialectology will be illustrated by applying the MLN approach to Chinese dialect data. The MLN approach is chosen for two reasons. Firstly, it offers a character-based framework to study language history. It yields direct, clear-cut hypotheses regarding the history of individual words and may thus be directly compared to alternative proposals. Secondly, the implementation of the MLN is freely available and scholars can easily replicate and further explore the results of the analysis.<sup>6</sup>

#### 3.1 Materials

##### 3.1.1 Data

The data used for this study was originally employed in a study by Ben Hamed and Wang (2006) and kindly provided by the second author. In this study, the authors investigated the suitability of lexicostatistical data for Chinese dialect classification by testing various algorithms, both for the reconstruction of phylogenetic trees and networks. The data consists of lexicostatistical wordlists for 23 Chinese dialect varieties and covers the seven traditional dialect groups. Each of the wordlists contains translations for 200 basic concepts (following Swadesh 1952 and 1955). Due to cases of synonymy where more than one translation or an optional variant was available for the same concept, the total number of words in the dataset amounts to 5349. All words were further coded for *cognacy* by assigning words which go back to the same ancestor form to a common *cognate set*. The original data offers different variants of cognate coding. For this study, a strict coding variant was chosen. This means that words are only assigned to the same cognate set if they match *completely*. This is a simplifying assumption, since—due to the frequency of compounding in Chinese—Chinese dialect words frequently exhibit *partial cognate relations* in which words contain cognate morphemes without matching completely. This is illustrated in Table 3, where translations of the concept ‘moon’ in four dialects share the common morpheme *yuè* 月 ‘moon’ (Middle Chinese \**ɲjot*) but use different (or no) suffixes (data taken from Hóu 2004). In order to handle these cases of partial cognacy adequately, specific models allowing for *multiple character states* are necessary (Nunn 2011: 59–60). Since the MLN approach currently only handles *binary character states* (i.e. presence vs. absence), the words need to be assigned to three different cognate sets in the analysis, as indicated by the different numbers for each cognate set in the table. This coding procedure yields a total of 1513 cognate

TABLE 3 *Partial cognacy in Chinese dialect words for ‘moon’ and resulting binary presence-absence-matrix*

Variety	Form	Character	Cognate-Set	Presence-Absence-Matrix		
				1 (月)	2 (月光)	3 (月亮)
Fúzhōu 福州	ɲuoʔ <sup>5</sup>	月	1	1	0	0
Měixiàn 梅縣	ɲiat <sup>5</sup> kuoŋ <sup>44</sup>	月光	2	0	1	0
Shànghǎi 上海	fiyɿ <sup>11</sup> lia <sup>23</sup>	月亮	3	0	0	1
Běijīng 北京	ye <sup>51</sup> liɑŋ <sup>1</sup>	月亮	3	0	0	1

6 For a coding example on how the MLN method can be applied to Chinese dialect data, see: <https://gist.github.com/LinguList/7481097>.

sets distributed over the 23 dialect varieties. However, since 922 of the cognate sets occur only once in the data and are thus not informative for the MLN approach, only 591 cognate sets were left for the analysis. The full dataset as it was used in this study is provided in Supplementary Material 11

### 3.1.2 Reference Trees

Using a reference tree is crucial for the application of the MLN approach. The reference tree represents the basic hypothesis regarding the vertical direction of character evolution and serves as the “genetic backbone.” Given the multitude of different proposals for Chinese dialect classification, it is useful to test several reference trees instead of employing only a single one when applying the MLN approach. For this study, a total of seven different reference trees was prepared. Three reference trees reflect independent hypotheses regarding Chinese dialect classification, as proposed in the literature. Three trees were reconstructed with the help of automatic methods for phylogenetic reconstruction. Additionally, a *random tree* was used to test the impact of arbitrariness on the MLN approach. All numbers reported for the random tree reflect the average results of 50 trials.

The first independent hypothesis is Laurent Sagart’s *Arbre des Dialectes Chinois* (Sagart 2011, personal communication). Based on distinct innovations for each split in the family tree, the Mǐn dialects were the first to split off, followed by Hakka and Yuè, which form a distinct subgroup. Originally, Sagart’s proposal contains not only the seven major dialect groups but also Wǎxiāng 瓦鄉 and Càijiā 蔡家, two archaic varieties which Sagart assumes to be the first varieties to split off the Sinitic branch of Sino-Tibetan. The second independent hypothesis is Jerry Norman’s *Southern Chinese Hypothesis* (Norman 1988: 210–214) which divides Chinese dialects into a northern, a central, and a southern zone (including Hakka, Yuè, and Mǐn), based on a small set of lexical and phonetic features. The third hypothesis is the *Hànyǔ Fāngyán Shùxíngtú* 漢語方言樹形圖 (“Tree chart of Chinese dialects”) by Yóu Rǔjié 游汝杰 (Yóu 1992: 91–106). This hypothesis is based on a thorough investigation of known population movements and the identification of dialect varieties with the respective ancient populations. Among others, it places Wú and Mǐn as close neighbors, as well as Gàn and Hakka. The three hypotheses, Sagart’s *Arbre*, Norman’s *Southern Chinese*, and Yóu’s *Shùxíngtú* are all illustrated in Figure 7. All hypotheses show only the subgrouping for the major dialect groups of Chinese. The finer groupings were taken from a maximum parsimony analysis of the full dataset described in the study by Ben Hamed and Wang (2006: 46).

Automatic reference trees were reconstructed using two popular distance-based methods for tree reconstruction, namely the *Unweighted Pair Group Method with Arithmetic Mean* (UPGMA, Sokal and Michener 1958) and *Neighbor-Joining* (Saitou and Nei 1987), and *maximum parsimony* as a popular character-based method. For the distance-based methods, the calculations were repeated for this study, using the UPGMA and

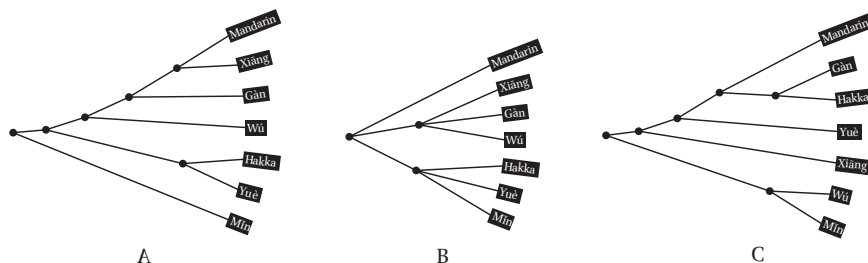


FIGURE 7 The three independently proposed hypotheses used as reference trees, *Arbre* (A), *Southern Chinese* (B), and *Shùxíngtú* (C).

*Neighbor-Joining* implementations provided by the LingPy library (List et al. 2015). For the character-based parsimony analysis, the results were taken from the study by Ben Hamed and Wang (2006: 47). Reference trees with full resolution, including all 23 dialect varieties for all automatic approaches and the independent hypotheses are given in Supplementary Material III.

### 3.2 *Methods*

The data was analyzed using the default settings of the most recent version of the MLN approach as implemented in LingPy (<http://lingpy.org>, Version 2.4.1-alpha, List et al. 2015). In the default settings, five different gain-loss models are tested. The models differ regarding the ratio between the penalties for gain and loss events, ranging from 3:1 over 2:1 to 1:1. The higher the penalty for gain events in comparison to loss events, the stronger the tendency of the model to favor vertical over horizontal transmission. Hence, the 3:1 model is the most conservative one, and the 1:1 model the most innovative. The MLN method also contains a specific parameter which handles the amount of parallel evolution allowed during model evaluation. According to the default settings, this parameter was set to 0.05, allowing ancestral vocabulary distributions to grow by 5% without affecting the choice of the best model. Using these settings, the MLN approach was applied to all seven reference trees. Scripts that reproduce the analyses described in this study, along with a computer-readable version of the data and the reference trees are given in Supplementary Material I (C and D).

### 3.3 *Accuracy of the MLN Approach*

It is important to get a clearer picture of the general accuracy of the MLN approach. In List et al. (2014a), this was done by testing to which degree known borrowings in an Indo-European dataset were readily identified by the method. In this study, the MLN approach identified 72% of all known borrowings correctly. Since we lack a list of known borrowings for the Chinese dialects, we cannot apply this test to the current dataset. However, there is a simple but effective method that may provide similar insights into the power of the MLN approach. This method is based on the “seeding” of fake borrowings (Dessimoz et al. 2008). Before analyzing the data, a certain amount of dialect pairs is randomly selected with one dialect serving as a donor and one as a recipient. A certain amount of words is then transferred from each donor to each recipient, thereby replacing the original word form in the given meaning slot.

Here, this method was applied to varying numbers of donor-recipient pairs (3, 6, 9, 12, and 15 pairs) with a varying borrowing rate  $\leq 25\%$ .<sup>7</sup> For each number of donor-recipient pairs, 50 different datasets with fake borrowings were produced and analyzed using the reference trees of the three independent

TABLE 4 *Performance on fake borrowing detection (see text)*

Borrowing Degree	<i>Arbre</i>	<i>Southern Chinese</i>	<i>Shùxíngtú</i>	<i>Neighbor-joining</i>	UPGMA	RANDOM
3	0.76	0.77	0.75	0.68	0.67	0.66
6	0.77	0.76	0.76	0.67	0.63	0.62
9	0.75	0.73	0.75	0.63	0.60	0.62
12	0.72	0.70	0.73	0.65	0.60	0.60
15	0.69	0.69	0.66	0.63	0.60	0.58

<sup>7</sup> The borrowing rate was chosen in such a way that it cannot exceed 25%, but it can be much less.

hypotheses, the RANDOM trees, and two automatic methods for tree reconstruction, namely UPGMA and *Neighbor-Joining*. Note that both automatic methods were applied to the data *after* the borrowings were introduced. The automatic analyses were thus forced to deal with the artificial reticulation in the data.

The averaged results for the tests are given in Table 5. For lower degrees of borrowing (up to 9 dialect pairs), all independent hypotheses perform fairly well on the borrowing-detection task. With higher degrees of borrowing, however, performance decreases drastically, approximating the results of the RANDOM analyses. This is not surprising, since, with higher degrees of borrowing involving more and more dialects, cognate sets will have reflexes in more and more dialects and start to resemble the inherited items. The MLN method thus successively loses its most important evidence: the patchiness of phyletic patterns. In contrast to the more or less satisfying performance of the analyses based on the independent hypotheses, the performance of the analyses based on automatic reference trees is alarmingly bad, showing only minor (if any) improvements compared to the RANDOM analysis. The reason for the bad performance of the automatic methods is that they are based on the ‘infected’ data. The introduction of artificial borrowings reduces and confuses the phylogenetic signal. A phylogenetic signal, however, is needed to guarantee that borrowings can be readily detected. As a result, the phylogenetic reconstruction methods fail to detect the vertical signal in the data, and the resulting reference trees are not suitable to detect horizontal signals with the help of the MLN approach. This shows the importance of selecting suitable reference trees when analyzing data with help of minimal lateral networks. It also shows how important it is for historical linguistics to carry out closer investigations on the borrowability of linguistic features. Scripts for the replication of this analysis are provided in Supplementary Material 1 (B).

### 3.4 Results

#### 3.4.1 General Results

Table 4 shows the general results of the analysis for all seven hypotheses.<sup>8</sup> As can be seen from the table, the 2:1 model shows the best fit with the VSD criterion in all analyses. The results for the six regular hypotheses do not differ much with respect to the number of inferred origins (gain events) per cognate set, ranging from 1.41 (*Arbre*) to 1.50 (UPGMA) gain events per cognate set on average. The same holds for the proportion of cognate sets which cannot be readily explained with help of the reference tree (column ‘Overall’ in the table). Here, the *Arbre* requires the lowest amount of horizontal evolution (34% of the cognate sets cannot be explained by vertical inheritance alone), and the *Shùxíngtú* the largest (39%). In contrast to the minor differences between the six regular hypotheses, the RANDOM analysis (based on

TABLE 5 General results of the analyses

Hypothesis	Model	Origins	Borrowings <sup>9</sup>			
			Overall	Yakhontov	Swadesh-100	Swadesh-200
<i>Arbre</i>	2:1	1.41	0.34	0.26	0.34	0.37
<i>Southern Chinese</i>	2:1	1.47	0.36	0.31	0.37	0.38
<i>Shùxíngtú</i>	2:1	1.49	0.39	0.36	0.36	0.43
UPGMA	2:1	1.50	0.38	0.30	0.36	0.42
<i>Neighbor-joining</i>	2:1	1.43	0.35	0.26	0.34	0.39
<i>Parsimony</i>	2:1	1.46	0.37	0.29	0.36	0.40
RANDOM	2:1	1.96	0.59	0.57	0.59	0.63

8 Scripts that replicate the analyses discussed in this section are provided in Supplementary Material 1 (C).



the averaged scores of 50 trials) shows an exceptionally high amount of 1.91 gain events and a large amount of inferred borrowing events. This shows the importance of choosing a good reference tree when applying the MLN approach.

Note that the differences in terms of origins and percentage of horizontal transfer events do not necessarily qualify or disqualify a given hypothesis. They just show how good a given hypothesis *explains* the data on which it is tested. A hypothesis which reveals that large amounts of the data cannot be explained by vertical transmission alone may be more realistic than one that reduces horizontal transmission to a minimum. If horizontal transmission is a genuine aspect of the language family under investigation, the former hypothesis may come much closer to the real history than the latter. In order to explore the different hypotheses in more detail, it is therefore important to take a closer look at the direct inferences of the hypotheses. Apart from the overall amounts of inferred horizontal events displayed in Table 4, the results for specific subsets are also given in the table. The column ‘Yakhontov’ lists the amount of inferred borrowings for all cognate sets which denote concepts belonging to Sergei Yakhontov’s list of the 35 most stable basic vocabulary items (Starostin 1995). The column ‘Swadesh-100’ lists the results for those cognate sets which appear on Morris Swadesh’s (1909–1967) list of the 100 most stable basic vocabulary items (Swadesh 1955) but not on Yakhontov’s list. The column ‘Swadesh-200’ shows the results for the rest of the 200 items which neither belong to Yakhontov’s nor Swadesh’s list. Assuming that the likelihood of a concept to be borrowed is the lowest for Yakhontov’s list, followed by Swadesh’s 100-item list, it is interesting to compare how the proportions of inferred horizontal events are distributed over the sublists. Here, the *Arbre* hypothesis yields the closest reflection of the sublist structure, proposing the smallest amount of horizontal transfer events for Yakhontov’s 35-item list, followed by Swadesh’s 100-item list, and the remainder of events in a third group. Interestingly, the sublist structure is also reflected in the RANDOM analysis, although the differences between the lists are less apparent.

Figure 8 shows the minimal horizontal network reconstructed from the *Shùxíngtú*.<sup>9</sup> The three heaviest edges in the MLN are between Méixiàn 梅縣 and Guǎngzhōu 廣州 (9 cognate sets), between the common ancestor of Shànghǎi 上海 and Sūzhōu 蘇州 and the common ancestor of Mǐn, Gàn, Hakka, Yuè, and Xiāng (7 cognate sets), and between Chángshā 長沙 and Mandarin (6 cognate sets). In Table 6, the three heaviest horizontal edges as inferred by the independent hypotheses are displayed, along with the forms that cause reticulation. The forms are further marked, depending on the sublists to which they belong. Concepts which appear on Yakhontov’s list are shaded in black, and concepts appearing on Swadesh’s 100-item list are shaded in gray. Numbers (shaded in dark gray on white font) reflect the Leipzig-Jakarta Index (Tadmor 2009) of conceptual items which are supposed to be very stable and highly resistant to borrowing (the smaller the number, the higher the supposed stability and resistance). As can be seen by scanning the data in the table, the numbers reported above find their reflection also in the three heaviest edges of the MLNs, with the *Arbre* involving less stable concepts, compared to the *Southern Chinese Hypothesis* or the *Shùxíngtú*. This does not only hold for the rankings proposed by Yakhontov and Swadesh, but also for the Leipzig-Jakarta Index (which is based on empirical data). This suggests that the *Arbre* mirrors the vertical history of the Chinese dialects more closely than the alternative hypotheses.

Whether the links inferred by the methods reflect real reticulation heavily depends on the reference tree. The nine horizontal edges which the *Shùxíngtú* infers for Guǎngzhōu and Méixiàn, for example, may as well reflect a closer genetic relationship between Hakka and Yuè (as proposed by the *Arbre* and the *Southern Chinese Hypothesis*). However, since the *Shùxíngtú* places Yuè and Hakka in separate

9 Plots for the minimal lateral networks of all other hypotheses are given in Supplementary Material V.

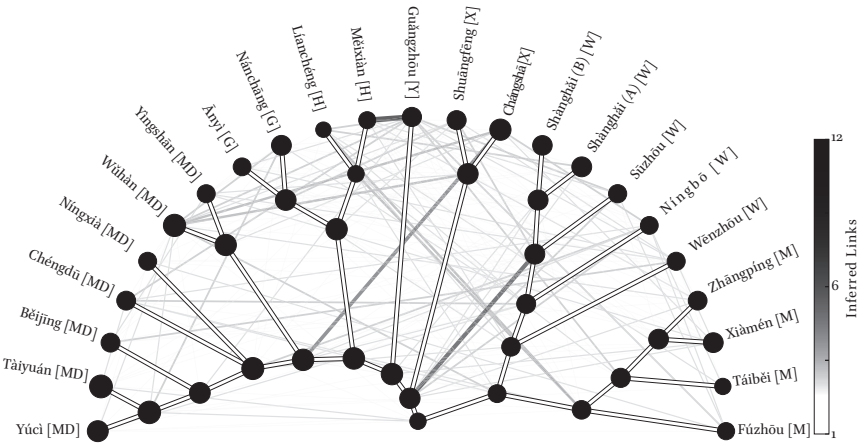


FIGURE 8 MLN for the Shùxíngtú hypothesis. Squared brackets behind the dialect names refer to the major dialect groups (G: Gà, MD: Mandarin, M: Mǐn, H: Hakka, W: Wú, X: Xiāng, Y: Yuè).

TABLE 6 The three heaviest edges inferred for the independent hypotheses. Background color of the items marks to which traditional “concept list” the items belong, with black indicating the list of 35 items by Yakhontov, gray indicating concepts occurring in Swadesh’s 100-item list, and the remainder of the 200-item list being left unmarked. Numbers (shaded in dark gray on white font) indicate the ranking according to the Leipzig-Jakarta index (Tadmor 2009) of highly stable conceptual items.

Hypothesis	Node A	Node B	Weight	Forms
Arbre	Xiàmén	non-Mǐn dialects	4	轉 ‘turn’, 拉 ‘pull’, 殺 ‘kill’, 擦 ‘rub’
	Níngbō	Mandarin (without Yíngshān, Wúhàn)	4	眼 ‘eye’ 83, 老的 ‘old’ 74, 咋 ‘how’, 舌 ‘tongue’ 6
	Gūangzhōu	Sūzhōu	4	男人 ‘husband’, 游泳 ‘swim’, 女人 ‘wife’, 如果 ‘if’
Southern Chinese	Wēnzhōu	Mǐn, Yuè, Hakka	7	卵 ‘egg’ 52, 溺水 ‘swim’ 42, 烏 ‘black’ 17, 驚 ‘fear’, 葉 ‘leaf’ 64, 腸 ‘guts’
	Chángshā	Mandarin	5	那個 ‘that’, 那裡 ‘there’, 鼻子 ‘nose’ 2, 女的 ‘woman’, 他們 ‘they’
	Guǎngzhōu	Wúhàn	5	男人 ‘husband’, 夜晚 ‘night’ 2, 游泳 ‘swim’, 爸爸 ‘father’, 扯 ‘pull’
Shùxíngtú	Měixiàn	Guǎngzhōu	9	嘍 ‘breathe’, 生 ‘live (alive)’, 生果 ‘fruit’, 唔 ‘not’ 56, 熟頭 ‘sun’, 掰 ‘split’, 名 ‘name’ 15, 利 ‘sharp’, 啱 ‘in’ 98
	Wú (without Níngbō, Wēnzhōu)	Mandarin, Gà, Yuè, Xiāng, Hakka	7	葉子 ‘leaf’ 64, 眼睛 ‘eye’ 83, 拿 ‘hold-take’, 瞞 ‘lie’, 嘔 ‘vomit’, 翅膀 ‘wing’ 17, 聞 ‘smell’
	Chángshā	Mandarin	6	那個 ‘that’, 那裡 ‘there’, 鼻子 ‘nose’ 2, 女的 ‘woman’, 游泳 ‘swim’, 他們 ‘they’

branches, it cannot explain the nine forms shared by the two dialects as the result of common inheritance and therefore has to explain them as horizontal transfer events.

3.4.2 Inferred Ancestral States and the Plausibility of Reference Trees

Since the MLN approach so heavily depends on the selection of the ‘correct’ reference tree, it is unsuitable to use the method directly to assess which of a given set of reference trees is the most plausible one. Without further external evidence, minimal lateral networks do not tell us whether a given hypothesis is wrong, or whether a given hypothesis is less suitable than another one. Nevertheless, minimal lateral



networks display the *consequences* of classification hypotheses in terms of distinct historical scenarios of character evolution, and these consequences can be evaluated with the help of external evidence.

Figure 9 shows an evolutionary scenario for the development of translations for ‘sun’ (*ritóu* 日頭 and *tàiyáng* 太陽) in the Chinese dialects as proposed by the MLN method. This scenario is far from being correct, since it proposes that the form *tàiyáng* 太陽 had developed in the common ancestor of Wú and Mandarin, although the Shànghǎi [tʰa<sup>33</sup>hiä<sup>44</sup>] 太陽 dialect’s form is a recent borrowing from Standard Chinese (Qian 2007: 24–27). Nevertheless, the scenario is better than many other possible scenarios, since it correctly places only the cognate set *ritóu* 日頭 in the ancestor languages of all dialects in the sample, rather than the late Mandarin innovation *tàiyáng* 太陽. Since the MLN approach regularly yields these scenarios for all characters in the data, we can test the plausibility of different classification hypotheses by checking how well the scenarios match with external evidence.

As a simple way to do so, it was examined which cognate sets are correctly traced back to the root of the reference trees. As external evidence, an Old Chinese wordlist, which was provided along with the original data used in Ben Hamed and Wang (2006), was used. Since not all word forms of this word list have reflexes in the data, only 148 forms were used for testing, and all other semantic slots were excluded from the analysis. In addition to the original MLN analysis (henceforth labeled *BestModel*), a further analysis, in which no horizontal transfer events were allowed was carried out (henceforth labeled *LossOnly*). The results for the two analyses are given in Table 7. As can be seen from the table, the models which do *not* allow for horizontal transfer trace more cognate sets back to the root of the tree. In the case of Southern Chinese and *Shùxíngtú*, this results in a hypothetical proto-language that contains more than 200 words to express less than 150 concepts. Once horizontal transfer is allowed, the number of cognate sets traced back to the proto-language shrinks drastically. Although all *Loss Only* analyses show a larger proportion of cognate sets which are correctly traced back to the root, their *precision* is much lower than

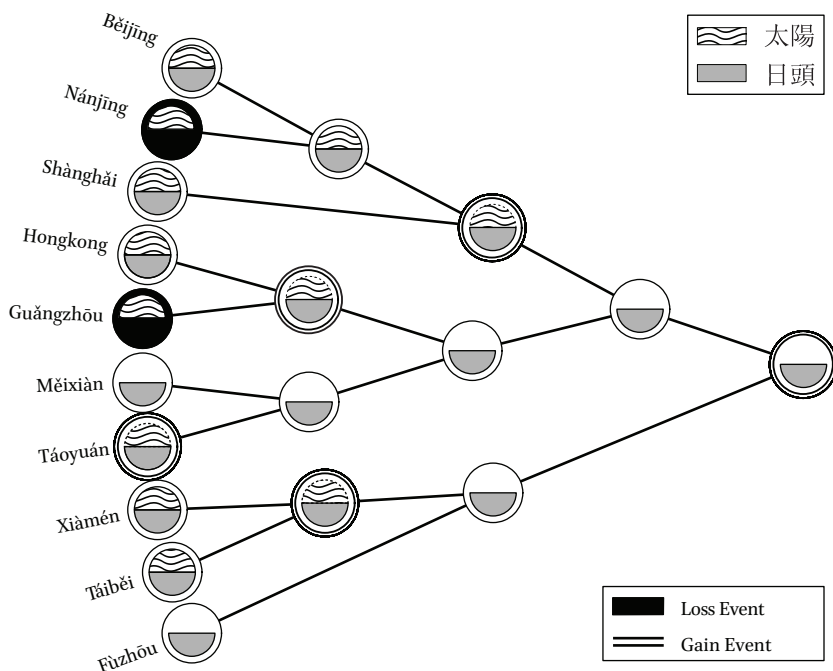


FIGURE 9 An evolutionary scenario for the development of ‘sun’ in some Chinese dialects.

TABLE 7     *Precision of ancestral state reconstruction for different models (see text)*

	<i>LossOnly</i>			<i>BestModel</i>		
	<i>Arbre</i>	<i>Southern Chinese</i>	<i>Shùxíngtú</i>	<i>Arbre</i>	<i>Southern Chinese</i>	<i>Shùxíngtú</i>
Proposed forms	148	225	215	107	132	119
Correct forms	94	114	115	81	83	86
Precision	0.64	0.52	0.54	0.76	0.63	0.72

that of the MLN analyses, since they produce a larger amount of false positives. This shows clearly, that the MLN approach yields more realistic reconstructions of ancestral states than pure loss-only models of character evolution.

Comparing the performance of the three hypotheses, the *Arbre* shows the highest precision, with 76% of all cognate sets traced back to the root occurring in the external wordlist of Old Chinese. The *Shùxíngtú* traces more cognate sets correctly back to the root, but it also produces a larger amount of false positives, resulting in a precision score of 72%. The *Southern Chinese Hypothesis* shows the lowest precision of all three independent classification proposals. Recalling that the *Arbre* also shows the best reflection of item stability as shown by the comparison of the proportion of inferred horizontal transfer events for the different sublists, one may preliminarily conclude that Sagart’s *Arbre des Dialectes Chinois* offers a more plausible genetic classification of the Chinese dialects than the *Southern Chinese Hypothesis* or the *Hànyǔ Fāngyán Shùxíngtú* hypothesis. Given that Sagart’s hypothesis is based on a carefully selected number of shared innovations, as opposed to the mix of innovations and retentions underlying Norman’s *Southern Chinese Hypothesis*, and the identification of known population movements with contemporary dialects underlying Yóu Rǔjié’s *Shùxíngtú*, this underlines the importance of shared innovations in language classification, which was already outlined by Karl Brugmann (1849–1919) in 1886 (Brugmann 1886 [1967]).

4     Discussion

What can we learn from this study? Are minimal lateral networks the key to reconstructing Chinese dialectal history, or are they just another fancy tool for visualization that explains problems away rather than facing them? The answer lies somewhere in the middle: minimal lateral networks surely do not solve all long-standing problems of Chinese dialectology at once. However, they are much more than the shiny images they yield in the end. Minimal lateral networks show the consequences of a hypothesis about the vertical paths of language history. In this respect, they are much more transparent than splits networks which only assess the tree-likeness of a given dataset. However, since minimal lateral networks are more transparent, they are also more easily exposed to criticism. Erroneous reference trees lead to erroneous inferences about horizontal language relations. If two closely related varieties are wrongly placed on distant branches of the reference tree, the MLN method will explain their genetic relations as a result of contact. This is a weakness of the method, but it is not a flaw.

But how good are the minimal lateral networks after all? When testing how well artificially seeded borrowings were detected by the MLN method, it became evident how important it is to have an *initial guess* regarding the “true” vertical history of the languages under investigation. This shows that minimal lateral networks are by no means a miracle cure against the horizontal forces of language history. They can help to identify the weak spots in our data, provided we have a good idea regarding the general vertical history of the languages being investigated. Yet, without an initial hypothesis regarding the genetic

history of a language family, they cannot tell us which items in our data are inherited and which are borrowed. Several factors contribute to this specific weakness of the minimal lateral network approach, the most important ones being (a) the underlying model of lexical change, and (b) the restriction of the model to use only tree-topologies to find conflicting signals.

The underlying model of lexical change is problematic, since it models lexical change as a simple process involving *cognate gain* and *cognate loss*. It is beyond question that such a model is far too simple to reflect lexical change realistically and accurately. In the current implementation of the MLN approach, for example, Shànghǎi [ɦyɿ<sup>21</sup> kuǎ̃<sup>23</sup>] 月光, Wēnzhōu [n̩y<sup>21</sup> kuɔ<sup>35</sup> vai<sup>13</sup>] 月光佛, and Běijīng [yɛ<sup>51</sup> lian̩<sup>1</sup>] 月亮 (all ‘moon’, data from Hóu 2004) are all treated as *etymologically unrelated words*. Since the MLN approach can only handle binary character states of character presence or character absence, their evolution is modeled independently of each other, although it is obvious that the word forms from Shànghǎi and Wēnzhōu are closer related to each other than to the form in Běijīng. It is straightforward to assume that the suffix [vai<sup>(2)13</sup>] was later added to the word in Wēnzhōu via the regular processes of suffixation (compare Wēnzhōu [n̩i<sup>21</sup> dʒu<sup>35</sup> vai<sup>13</sup>] 日頭佛 ‘sun’), and that the common ancestor of Wēnzhōu and Shànghǎi would denote ‘moon’ with a word of the form 月光, while the common ancestor of Wēnzhōu, Běijīng and Shànghǎi would not. What we are dealing with here are *oblique etymological relations* (List 2014: 41–44), or—in a more narrow sense—*partial cognate relations*, which are very frequent in the Sino-Tibetan languages. Binary gain-loss models of lexical change cannot handle these relations. In order to handle them, more complex models which allow for *multiple character states* and *character transition probabilities* are needed (Nunn 2011: 59–60). In biology, the use of multi-state models is quite common. It is, however, difficult to apply them to linguistic data, since most existing approaches have been specifically designed to deal with biological data, allowing only for a limited amount of possible character states or restricting transition probabilities to biologically meaningful values, which would not make sense in linguistic applications.

The restriction to use tree-topologies as the sole evidence for proposed horizontal language relations is another shortcoming of the of the MLN approach. Given that processes of horizontal transfer go along with areal proximity between language varieties, it seems tempting to take geographic information into account when dealing with questions of language contact. List et al. (2014b) illustrate how *minimal spatial networks* can be reconstructed from minimal lateral networks by taking areal information into account and drawing horizontal edges between language varieties on a geographic map. The study shows that minimal spatial networks add a useful perspective to the MLN approach and may help to reveal cases of close language contact due to areal proximity. However, since they only reflect contemporary language varieties, they can only reveal recent cases of language contact. This is because deeper relations, involving contact between ancestral languages which are represented as internal nodes in minimal lateral networks are not displayed. This reflects a general problem of using geographic information when trying to identify contact relations between languages: Given that we know that waves of migration may play a crucial role during language history, contemporary areal data often provides little evidence regarding deeper layers of contact, and our methods should carefully balance the different pieces of evidence on which they are based.

The obvious shortcomings involving the MLN approach show that the frequently mentioned parallels between biological evolution and language history cannot be strained *ad infinitum*. The application of biological methods in historical linguistics should always be based on a careful *adaptation* rather than a simple *transfer*. Nevertheless, even in their current, simplified form, minimal lateral networks show promising results. They are a useful starting point for further investigations in historical linguistics in general, and in Chinese dialectology in particular and it seems worthwhile to develop and test them further.

## Supplementary Material

The supplementary material accompanying this study consists of four parts, including the source code to replicate all analyses discussed in this study (I), a PDF file listing the dataset in human-readable format (II), a PDF file containing the plots of all reference trees used in this study (III), and a PDF file containing the plots of MLN analyses for all reference trees (IV). The supplementary material can be found at: <http://dx.doi.org/10.5281/zenodo.16760>.

## Acknowledgements

This research was partially supported by the DFG grant 261553824 (<http://gepris.dfg.de/gepris/projekt/261553824>). I would like to thank Wáng Fēng for being so kind as to provide the data on Chinese dialects that was used in this study, and the two anonymous reviewers for their helpful comments.

## References

- Aikhenvald, Alexandra Y. 2007. Grammars in contact. A cross-linguistic perspective. In: Aikhenvald, Alexandra Y., Dixon, R. M. V. (eds.), *Grammars in contact*. Oxford: Oxford University Press, 1–66.
- Bandelt, H. J., Dress, A. W. 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution* 1 (3): 242–252.
- Baxter, William H. 2006. Mandarin dialect phylogeny. *Cahiers de Linguistique – Asie Orientale* 35 (1): 71–114.
- Běijīng Dàxué. 北京大學 (ed.). 1962 [1989]. 《漢語方音字匯》 [Chinese dialect character pronunciation list]. 北京：文字改革出版社.
- Ben Hamed, Mahé. 2005. Neighbour-Nets portray the Chinese dialect continuum and the linguistic legacy of China's demic history. *Proceedings of the Royal Society London, Series B* 272 (1567): 1015–1022.
- Ben Hamed, Mahé, Wang, Feng. 2006. Stuck in the forest: Trees, networks and Chinese dialects. *Diachronica* 23: 29–60.
- Branner, David P. 2000. *Problems in comparative Chinese dialectology. The classification of Min and Hakka*. Berlin and New York: Mouton de Gruyter.
- Brugmann, Karl. 1886 [1967]. *Einleitung und Lautlehre: vergleichende Laut-, Stammbildungs- und Flexionslehre der indogermanischen Sprachen*. In: *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen*. Berlin, Leipzig: Walter de Gruyter.
- Bryant, David, Moulton, Vincent. 2004. Neighbor-Net. An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21 (2): 255–265.
- Cohen, Ofir, Pupko, Tal. 2010. Inference and characterization of horizontally transferred gene families using stochastic mapping. *Molecular Biology and Evolution* 27 (3): 703–713.
- Dagan, Tal, Martin, William. 2007. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proceedings of the National Academy of Sciences U.S.A.* 104 (3): 870–875.
- Dagan, Tal, Artzy-Randrup, Y., Martin, William. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proceedings of the National Academy of Sciences U.S.A.* 105 (29): 10039–10044.
- Dessimoz, Christophe, Margadant, Daniel and Gonnet, Gaston H. 2008. DLIGHT – Lateral gene transfer detection using pairwise evolutionary distances in a statistical framework. In: Vingron, M., Won, L. (eds.). *Research in Computational Molecular Biology*. Berlin and Heidelberg: Springer, 315–330.

- Dunn, Michael, Terrill, Angela, Reesink, Ger, Foley, Robert A., Levinson, Stephen C. 2005: Structural phylogenetics and the reconstruction of ancient language history. *Science* 309 (5743): 2072–2075.
- Hóu, Jīngyī. 侯精一 (ed.). 2004. 《現代漢語方言音庫》 [Phonological database of Chinese dialects]. 上海：上海教育出版社.
- Huson, Daniel H. 1998. “SplitsTree. Analyzing and visualizing evolutionary data”. *Bioinformatics* 14: 68–73.
- Huson, Daniel H. Scornavacca, Celine. 2011. A survey of combinatorial methods for phylogenetic networks. *Genome Biology and Evolution* 3: 23–35.
- Koonin, Eugene V., Makarova, Kira S., Aravind, L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review Microbiology* 55: 709–742.
- Kurpaska, Maria. 2010. *Chinese language(s). A look through the prism of “The Great Dictionary of Modern Chinese Dialects”*. Berlin and New York: de Gruyter.
- Lǐ, Xiǎofán. 李凡. 2005. 〈漢語方言分區方法再認識〉 [Reevaluating the classification of the Chinese dialects]. 《方言》 4: 356–363.
- List, Johann-Mattis, Moran, Steven. 2013. An open source toolkit for quantitative historical linguistics. In: *Proceedings of the ACL 2013 System Demonstrations* (Sofia, 08/04–08/09/2013). Stroudsburg: Association for Computational Linguistics, 13–18.
- List, Johann-Mattis, Moran, Steven, Bouda, Peter, Dellert, Johannes, Rama, Taraka, Forkel, Robert. 2015: LingPy. Python library for automatic tasks in historical linguistics. Version 2.4.1-alpha (Uploaded on 2015-03-16). URL: <http://lingpy.org>. DOI (Zenodo): <http://dx.doi.org/10.5281/zenodo.16093>.
- List, Johann-Mattis, Nelson-Sathi, Shijulal, Geisler, Hans, Martin, William. 2014a. Networks of lexical borrowing and lateral gene transfer in language and genome evolution. *BioEssays* 36 (2): 141–150.
- List, Johann-Mattis, Nelson-Sathi, Shiju, Martin, William, Geisler, Hans. 2014b. Using phylogenetic networks to model Chinese dialect history. *Language Dynamics and Change* 4 (2): 222–252.
- List, Johann-Mattis. 2014. *Sequence comparison in historical linguistics*. Düsseldorf: Düsseldorf University Press.
- McMahon, April, Heggarty, Paul, McMahon, Robert, Slaska, Nathalia. 2005. Swadesh sublists and the benefits of borrowing: An Andean case study. *Transactions of the Philological Society* 103: 147–170.
- Mirkin, Boris G., Fenner, Trevor I., Galperin, Michael Y., Koonin, Eugene V. 2003. Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evolutionary Biology* 3 (2).
- Nakhleh, Luay, Ringe, Don, Warnow, Tandy. 2005. Perfect Phylogenetic Networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81 (2): 382–420.
- Nelson-Sathi, Shijulal, List, Johann-Mattis, Geisler, Hans, Fangerau, Heiner, Gray, Russell D., Martin, William, Dagan, Tal. 2011. Networks uncover hidden lexical borrowing in Indo-European language evolution. *Proceedings of the Royal Society London, Series B* 278 (1713): 1794–1803.
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Norman, Jerry. 2003. The Chinese dialects. Phonology. In: Thurgood, Graham and LaPolla, Randy (eds.), *The Sino-Tibetan languages*. London and New York: Routledge, 72–83.
- Norman, Jerry and Coblin, Weldon S. 1995. A new approach to Chinese historical linguistics. *Journal of the American Oriental Society* 115 (4): 576–584.
- Nunn, Charles L. 2011. *The comparative approach in evolutionary anthropology and biology*. Chicago and London: University of Chicago Press.
- Qián Nǎiróng. 錢乃榮. 2007. 《上海方言》 [The Shànghǎi dialect]. 上海：文匯出版社.
- Sagart, Laurent. 2001. Vestiges of Archaic Chinese derivational affixes in modern Chinese dialects. In: H Chappell, Hilary (ed.), *Sinitic grammar. Synchronic and diachronic perspectives*. Oxford: Oxford University Press, 123–142.



- Sagart, Laurent. 2011. Classifying Chinese dialects/Sinitic languages on shared innovations. Paper, presented at the *Séminaire Sino-Tibétain du CRLAO* (Paris, 03/28/2011).
- Saitou, Naruya and Nei, Masatoshi. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4 (4): 406–425.
- Schleicher, August. 1853. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur* 3: 786–787.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.
- Sokal, Robert R., Michener, Charles D. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin* 28: 1409–1438.
- Southworth, Franklin C. 1964. Family-tree diagrams. *Language* 40 (4): 557–565.
- Starostin, Sergej A. 1995. Old Chinese vocabulary: A historical perspective. In: Wang, William S.-Y. (ed.), *The ancestry of the Chinese language*. Berkeley: University of California Press, 225–251.
- Swadesh, Morris. 1952. Lexico-statistic dating of prehistoric ethnic contacts. With special reference to North American Indians and Eskimos. *Proceedings of the American Philological Society* 96 (4): 452–463.
- Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21 (2): 121–137.
- Tadmor, Uri. 2008. Loanwords in the world's languages. In: Haspelmath, Martin, Tadmor, Uri (eds.), *Loanwords in the world's languages*. Berlin and New York: de Gruyter, 55–75.
- Wáng, Hóngjūn. 王洪君. 2009. 〈兼顧演變、推平和層次的漢語方言歷史關係模型〉 [A historical relation model of Chinese dialects with multiple perspectives of evolution, level and stratum]. *《方言》* 3: 204–218.
- Wang, William S.-Y. 1997. Languages or dialects? *The CUNY Journal of Humanities* 1: 54–62.
- Weinreich, Uriel. 1953. *Languages in contact*. The Hague and Paris: Mouton.
- Wurm, S. A., Liu, Yongquan (eds.). 1987. *《中國語言地圖集》* [Language atlas of China]. Hong Kong: Longman Group.
- Yan, Margaret Mian. 2006. *Introduction to Chinese dialectology*. München: LINCOM Europa.
- Yóu, Rǔjiá. 游汝杰. 1992. *《漢語方言學導論》* [Chinese dialectology]. 上海：上海教育 出版社.
- Zavjalova, O. I. 1996. *Диалекты китайского языка* [The Chinese dialects]. Москва: Научная Книга.
- Zhōngguó Shèhuì Kēxuéyuàn Yǔyán Yánjiūsuǒ. 中國社會科學院語言研究所 (ed.). 1955 [1981]. *《方言調查字表》* [List of characters for research on Chinese dialects]. 北京：商務印書館.

## 從演化網絡的角度來研究漢語方言史——機遇與挑戰

游函

法國國家科學研究院

### 提要

目前，學界對漢語方言歷史的了解還不夠完整。語言學家運用各種傳統方法與標準，試圖從假設的共同祖語來擬定聲調或音段的發展，從而界定主要的方言區。但是，有關聲調、音段的發展細節，學界尚遠未達成共識。學者們意見不統一的主要原因在於語言接觸在漢語方言歷史上所扮演的角色。漢語方言的情況跟其他語言不一樣：它們從來沒有真正分開過，也從來沒有獨立地發展，而是不斷在相互影響中演變。因此，我們很難確定方言之間共同特征是從祖語繼承而來的還是從其他方言中借過來的。最新的生物學演化網絡方法可以作為借鑒，這昭示出歷史語言學一個新的探索方向。這類方法運用縱向及橫向網絡以了解細菌的演化過程。語言學家開始運用這些方法來研究語言的發展，并已獲得可喜的初步成績。本文以漢語方言作為專題研究，運用演化網絡方法來分析漢語歷史語言學的問題。在具體操作上，我們精心挑選了200個核心詞彙，並在該基礎上建立了23種漢語方言的數據集。隨後，筆者針對漢語方言的各種分類理論作出了不同的比較和測試，從而判斷其普遍合理性。比較的結果表明，演化網絡方法對漢語方言史的定量和定性研究起到了舉足輕重的補充作用。為了充分發揮其潛力，我們需要建立更適合語言學研究的演化發展模型，例如加入地理信息等其他佐證，才能夠更全面、更細緻地描述漢語方言的歷史演變。

### 關鍵詞

漢語方言分類、漢語方言史、演化網絡方法、詞匯借用