


## Article

# Chinese Word Order in the Comparative Sino-Tibetan and Sociotypological Contexts

Chingduang Yurayong <sup>1,2,\*</sup>  and Erika Sandman <sup>1</sup>

<sup>1</sup> Department of Languages, Faculty of Arts, University of Helsinki, 00100 Helsinki, Finland; erika.sandman@helsinki.fi

<sup>2</sup> Research Institute for Languages and Cultures of Asia, Mahidol University, Nakhon Pathom 73170, Thailand

\* Correspondence: chingduang.yurayong@helsinki.fi

**Abstract:** The present study discusses typology and variation of word order patterns in nominal and verb structures across 20 Chinese languages and compares them with another 43 languages from the Sino-Tibetan family. The methods employed are internal and external historical reconstruction and correlation studies from linguistic typology and sociolinguistics. The results show that the head-final tendency is a baseline across the family, but individual languages differ by the degree of head-initial structures allowed in a language, leading to a hybrid word order profile. On the one hand, Chinese languages consistently manifest the head-final noun phrase structures, whereas head-initial deviants can be explained either internally through reanalysis or externally through contact. On the other hand, Chinese verb phrases have varied toward head-initial structures due to contact with verb-medial languages of Mainland Southeast Asia, before reinstalling the head-final structures as a consequence of contact with verb-final languages in North Asia. When extralinguistic factors are considered, the typological north-south divide of Chinese appears to be geographically consistent and gradable by the latitude of individual Chinese language communities, confirming the validity of a broader typological cline from north to south in Eastern Eurasia.

**Keywords:** Chinese; Sino-Tibetan; word order; variation; historical linguistics; linguistic typology; areal linguistics



**Citation:** Yurayong, Chingduang, and Erika Sandman. 2023. Chinese Word Order in the Comparative Sino-Tibetan and Sociotypological Contexts. *Languages* 8: 112. <https://doi.org/10.3390/languages8020112>

Academic Editors: Umberto Ansaldi and Pui Yiu Szeto

Received: 6 February 2023

Revised: 26 March 2023

Accepted: 5 April 2023

Published: 19 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Comparative studies of Chinese languages have become more widely practised in recent decades, particularly thanks to an increasing amount of information and data on understudied Chinese languages. Such valuable language descriptions supplement previous comparative works on Chinese dialectology (Yue-Hashimoto 1999; Li 2002; Cao 2008; Kurpaska 2010) and contribute to our better understanding of variation across Chinese languages. The advancement of such a diversity-based approach provides more evidence against the approach of “Universal Chinese Grammar” (Chao 1968, p. 13), which has been influential in Chinese linguistics for many decades (see also its criticism in Matthews 1999).

Such trends and advancement in the field of Chinese and Asian linguistics as well as linguistic typology in general have enabled further investigational attempts, which confirm an early observation by Hashimoto (1976) and strengthen the current consensus that variation in word order patterns of nominal and verb structures across Chinese languages largely results from language contact with non-Chinese languages. This emerging variation has been considered in a number of studies as evidence for areal diffusion in particular locations (Ansaldi 1999, 2010; Peyraube 2015; Szeto et al. 2018; Szeto 2019; Szeto and Yurayong 2021), particularly in Northwest China (Janhunen 2007; Sandman and Simon 2016) and South China (Li 2008; Matthews 2007; de Sousa 2015; Huang and Wu 2018; Szeto and Yurayong 2022).

Given the state of the art, the present study revisits earlier accounts of Chinese word order typology discussed in the context of the Sino-Tibetan family (Dryer 2003; LaPolla 2015). We aim to analyse diachronic changes of Sino-Tibetan word order based on synchronic data, naturally since diachronic data are only available for Chinese, Bodic and Burmo-Qiangic branches. For example, Nichols (1992) and Bickel (2015) have demonstrated that synchronic distributions of typological features can predictively provide information on diachronic trends and geographical clines as a directional implication for spread of languages and their features. Previously, variation and changes in the word order structure of Chinese have been discussed in a number of studies (e.g., Sun 1996; Xu 2006). However, Chinese has often been presented in publications as one name and one language possessing a single homogenous grammar contrasting with other more diverse Tibeto-Burman languages, leaving subtle differences present among individual Chinese languages unveiled. Unlike what has been usually done in previous studies, we adopt a diversity-linguistic approach which treats word order features of Chinese together with other Tibeto-Burman languages as equal datapoints without assumption or bias towards the hypothetical first level split of Proto-Sino-Tibetan into two major branches: Chinese vs. Tibeto-Burman (à Matisoff 2003; Sagart et al. 2019). The use of the term “Tibeto-Burman” in this study is therefore only for the purpose of referring to Sino-Tibetan languages which are not classified as belonging to the Chinese branch.

The aim of the current study is to tackle two main research questions, regarding the reconstruction of Proto-Chinese syntax and the role of variation and sociolinguistic factors in its change over time. By sociolinguistic factors, we refer to social factors related to language use in a broad sense, including population size, the number of L2 speakers and their speech communities. While these factors are usually not at the core of variationist sociolinguistics, they have received considerable attention in the work of sociolinguists interested in language change and language contact (e.g., Weinreich 1974; Thomason and Kaufman 1988). More recent works have also connected these factors to tendencies of typological change, including morphological complexity (e.g., Lupyan and Dale 2010; Sinnemäki and Di Garbo 2018). Although it is not possible to collect detailed sociolinguistic data on all the languages in our sample, the relevance of sociolinguistic factors for typological change have received so much attention in earlier research that we consider it necessary to bring out this issue as one of the factors which can explain our observations.

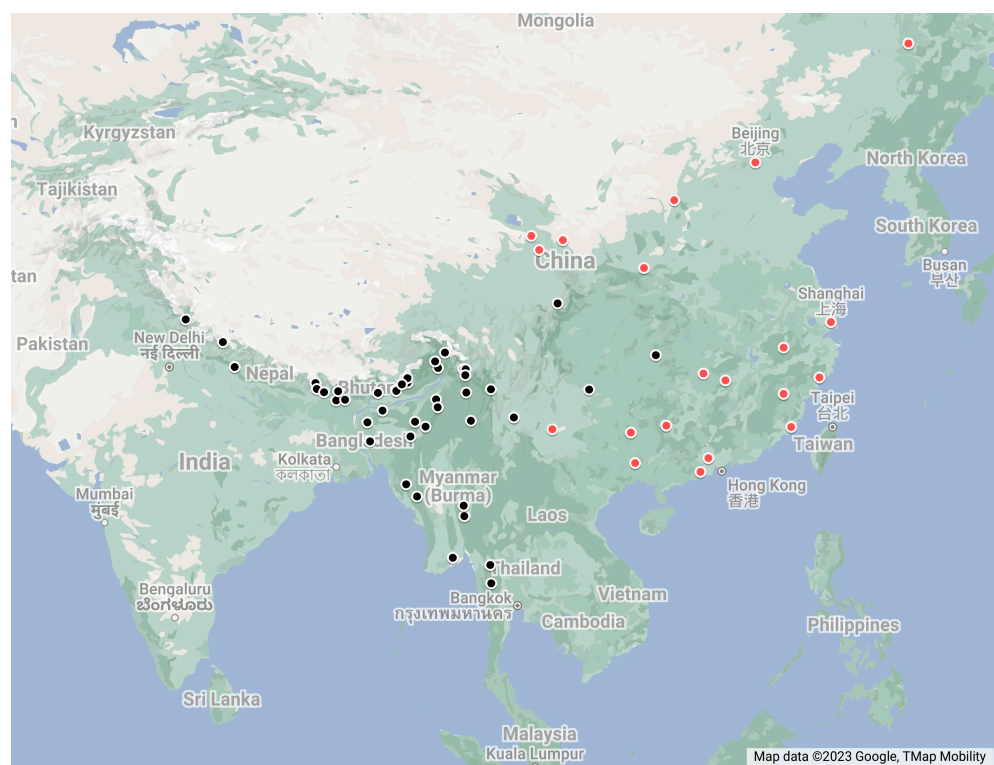
For the first research question, we explore diversity of word order structures across Chinese and the entire Sino-Tibetan language family through a variety sampling method developed by Miestamo et al. (2016). The principle is that one variety per one subbranch is taken in order to “display as much variety as possible in the linguistic realizations of the phenomena under investigation and to reveal even the rarest strategies or types of expression in the domain explored” (Miestamo et al. 2016, p. 234). The sample includes 20 Chinese and 43 Tibeto-Burman languages, geographically distributed across the Himalayas, their foothills in Southeast Asia and China as shown in Figure 1.

The data are collected from reference grammars and individual publications which provide synchronic (but sometimes also diachronic) descriptions and examples of individual word order features. Sociolinguistic data related to numbers of speakers and geographical locations of speech communities are also collected.

In terms of diachronic analysis to explain changes in word order typology of Sino-Tibetan languages, the current study uses a typological framework for both internal and external reconstruction methods. In practice, it means that we try to find an internal explanation first and will consider an external explanation only after the former reaches its limit, following a methodological practice of Thomason (2010). We challenge a holistic view that a protolanguage of any hitherto known language family must have been purely either the head-initial or head-final type without room for variation and hybridity.

As for the second research question, the data are quantitatively processed through coefficient of determination ( $R^2$ , Wright 1921) for testing correlations between intralinguistic (a structural aspect of language) and extralinguistic (social and geographical) variables

(see Section 3 for detailed description of the method). Concretely, we quantify variables by taking numbers of speakers and a geographical location (GIS coordinates) of any given speech community. The discussion pays special attention to a scenario of change initiated from sociolinguistically motivated free variation which further leads to the deviation of word order structure from the predominant head-final tendency of the Proto-Sino-Tibetan syntax, as is generally presented in the literature (see Dryer 2003; LaPolla 2015).



**Figure 1.** Geographical coverage of the data:  $n = 63$  (red = Chinese, black = Tibeto-Burman).

As a roadmap of the work structure, Section 2 provides relevant information and observations made in previous studies on word order and its related sociolinguistic factors in general and in Sino-Tibetan languages specifically. Section 3 describes data related matters and the method of analysis which is applied to the description of individual word order features in Section 4. The information and insights obtained from analysis of the data are used in the discussion of reconstruction and the explanation of directionality changes in word order from the Proto-Chinese and Proto-Sino-Tibetan stages, as shown in Section 5. Section 6 concludes the study with essential notes and suggestions for conducting further studies of the topic.

## 2. Background and Theoretical Framework

This section reviews significant observations of Chinese word order structure, as discussed in previous studies, and serves as a basis for building the research method for the present study (Section 3), as references for description of different word order features (Section 4) and as a theoretical framework for the discussion of reconstruction and the directionality of change in Chinese word order, motivated by multiple intralinguistic and extralinguistic factors (Section 5).

### 2.1. Typology and Directionality Change of Word Order

A nuclear element of syntax is labelled as “head” which interacts with its complements labelled as “dependent(s)”. A formalist approach assumes that a language has strictly consistent parameters for the order of head and its dependents (e.g., Jackendoff

1977; Lightfoot 1982; Chomsky and Lasnik 1993). However, natural languages are not always confined to such invariable parameters; instead, deviations can be observed, and the most we can say is that a language possesses a cross-category harmony governing tendencies of placing head before or after dependent(s) (e.g., Greenberg 1963; Hawkins 1983; Dryer 1992a). The present study will illustrate concrete evidence from Chinese and other Sino-Tibetan languages against the formalist idea of invariable parameters (as will be demonstrated in Sections 4 and 5).

Throughout the entire work, we adopt a dichotomy widely used in the literature on syntax to describe a structure and directionality where head precedes dependent(s) as “head-initial”, e.g., Vietnamese *năm nay* [year this] ‘this year’, whereas another structure where head follows dependent(s) as “head-final”, e.g., Mongolian *ene jil* [this year] ‘this year’ (see also alternative terminologies, e.g., “right” and “left branching” in Dryer 1992a). To give some examples from neighbouring languages of Chinese, Tai-Kadai, Hmong-Mien and Austroasiatic predominantly use head-initial nominal and verb structures, while languages of the Altaic type, Japonic, Koreanic, Tungusic, Mongolic and Turkic, consistently conform with head-final directionality (see numerous chapters on word order features in Dryer and Haspelmath 2013 for a worldwide survey).

Due to its binary characteristics, when a change of head directionality occurs in word order, it tends to be abrupt or is even sometimes considered catastrophic (Lightfoot 1999, p. 105). From the opposing viewpoint, the directionality change may not occur catastrophically over a night, but the result is a predictable outcome of competition among existing variants which are available and competing in the speakers’ linguistic resources at a certain point before one has won over the others (see the variationist approaches to language change in Kroch 1989; Santorini 1993; Pintzuk 2017). In the current study, we adopt the latter variationist approach and will show varying degrees of directionality change in individual phrase and clause structures in Sino-Tibetan languages (see Sections 4 and 5).

Regarding the predictability of changes in head directionality, we are always dealing with a finite set of possibilities, particularly when there are only binary options for directionality of word order, as discussed above. The attempt has been made to explain the directionality of word order in a typological framework by providing a relative chronology in which constituents and their related constructions shift their directionality faster. For instance, Greenbergian implicational universals (1963) have been applied in the discussion of hierarchy among different noun modification types in Hawkins (1983, p. 75), Dryer (1992b) and Croft (2003, p. 123), as shown in Table 1. The given constituent orders in each row represent word order profiles, six scenarios which are typologically expected in a language.

**Table 1.** The Prepositional Noun Modifier Hierarchy (modified from Hawkins 1983, p. 75; Dryer 1992b; and Croft 2003, p. 123): N = Noun; Quan = Quantifier; Dem = Demonstrative; Adj = Adjective; Poss = Possessor; and Rel = Relative clause.

Type	Head-Initial Structures	Head-Final Structures
1	NQuan & NDem & NAdj & NPoss & NRel	–
2	NDem & NAdj & NPoss & NRel	QuantN
3	NAdj & NPoss & NRel	QuantN & DemN
4	NPoss & NRel	QuantN & DemN & AdjN
5	NRel	QuantN & DemN & AdjN & PossN
6	–	QuantN & DemN & AdjN & PossN & RelN

When they are put in an ordered rank below, we can apply their hierarchical order to conduct a diachronic prediction of which types tend to shift their directionality earlier.

Quantifiers → Demonstratives → Adjectives → Possessors → Relative clauses

According to the proposed hierarchy, quantifiers being at the higher surface of noun phrase structure tend to shift their directionality first whereas relative clauses at a deeper structural level would be resistant and the last among the noun phrase constituents to shift their position. However, this postulation based on the noun phrase modification hierarchy is not necessarily valid universally across languages of the world (as will be shown in Section 5.3), but it can serve as a starting point for the prediction of syntactic change, which utilises a predictive power of the typological approach to language change (Greenberg 1957, p. 77; Jakobson 1958, p. 528). The Chinese and Sino-Tibetan data discussed in the present study will contribute further to the improvement of such a predictive tool.

When it comes to the extension of directionality changes, the effects may apply at different levels of syntactic structure in different rates. For instance, it has been discussed in diachronic syntax literature that subordinate clauses tend to be more conservative and therefore can preserve older patterns better than main clauses (Bybee 2001). This assumption is applicable, for example, to West Germanic languages in which main clauses have shifted to the verb-medial pattern (1a), while subordinate clauses still retain the Proto-Germanic verb-final pattern (1b).

#### Dutch (Indo-European)

- (1) a. *Ze spreken geen Chinees.*  
 3PL [speak.3PL no Chinese]  
 'They do not speak Chinese.'
- b. *Het is jammer dat ze geen Chinees spreken.*  
 it be.3sg pity QUOT 3PL [no Chinese speak.3PL]  
 'It is pity that they do not speak Chinese.'

A similar explanation has also been applied to trace the verb-initial structures inherited from Proto-Austroasiatic, which are retained in subordinate clauses of modern Austroasiatic languages, although they may have shifted to verb-medial or verb-final patterns due to contact and may vary in terms of how extensively the retention of the verb-initial structures still applies across the grammatical system (Jenny 2015, 2020). This assumption and related observations, which Chinese and other Sino-Tibetan languages provide evidence for (see Section 5.2), are useful also for the current study which conducts a syntactic reconstruction of Proto-Chinese and beyond by also paying attention to differences in the word order patterns of main vs. complex clauses.

## 2.2. Word Order Typology of Chinese and Sino-Tibetan Languages

In linguistic typology, Chinese and Sino-Tibetan languages are generally regarded as belonging to a head-final type of language, meaning that they tend to prefer word order with the head following dependents. This view is largely based on the observation that verb-final is the predominant syntactic pattern among the majority of Sino-Tibetan languages, which typologically entails correlations with other head-final structures (as discussed in Greenberg 1963; Dryer 1992a). However, individual Sino-Tibetan languages show varying degrees of deviation from this baseline head-final tendency, with Chinese exhibiting significantly more variation and admixture of head-final and head-initial structures. It has been argued that this phenomenon is best explained by language contact and areal diffusion (as examined by Dryer 2003; Chappell et al. 2007; Szeto and Yurayong 2021).

As illustrated in Table 2, the noticeably non-harmonious profile of word order structures in Chinese languages is evident particularly in verb phrases due to their predominant verb-medial basic word order, unlike noun phrases which are largely head-final.



**Table 2.** General tendencies of word order features of Chinese languages (based on [Chappell et al. 2007](#), p. 189).

Head-Final Structures	Head-Initial Structures
Adjective → Noun	Verb ← Direct Object
Possessor → Possessee	Auxiliary Verb ← Main Verb
Numeral → Classifier → Noun	Verb ← Adverbial complement
Demonstrative → Classifier → Noun	Adposition ← Noun phrase
Relative Clause → Noun	Complementiser ← Sentence
Degree adverb → Adjective	
Standard of comparison → Adjective	
Adverb → Verb	
Adpositional phrase → Verb	

Note that this is a generalisation which does not take into account the entire subtle variation observed across Chinese languages (for more data and analyses on variation across Chinese, see e.g., [Yue-Hashimoto 1999](#); [Szeto 2019](#), and examples in Section 4).

Previous discussions by [Dryer \(1992a, 2003\)](#) postulate that Chinese languages have more consistently preserved the original head-final structures in noun phrases, while verb phrases are affected by the assumed directionality change in basic word order from the verb-final to verb-medial pattern. However, other Sino-Tibetan languages, which retain the verb-final basic word order, also use the head-initial order for adjectival and quantity modifications as well as degree adverbs, all of which appear as head-final in Chinese languages. The situation is indeed fuzzier for other Sino-Tibetan languages as there are likewise a lot of hybrid phrase patterns (as discussed in [LaPolla 2015](#)), which makes it worth to compare them with Chinese in the present study, in order to shed light on the fact that the Sino-Tibetan syntax is far from being internally harmonious and homogenous. Many intralinguistic and extralinguistic factors have been responsible for head-dependent word order in the languages today (see further discussion in Section 5).

Coming to language contact issues, neighbouring languages to the north of Chinese, namely Tungusic, Mongolic and Turkic, consistently use head-final structures for all the features mentioned in Table 2 (as has been extensively discussed in [Szeto et al. 2018](#); [Szeto 2019](#); [Szeto and Yurayong 2021](#)). This is thought to have influenced the preference for a head-final tendency in northern Chinese languages, one of the components of a larger “Altaicisation” process occurring towards the northern zone of Chinese speaking areas, where language shift from a language of the Altaic type to Chinese has been continuously taking place during the past millennium ([Hashimoto 1976](#); [Janhunen 2007, 2012](#)). On the contrary, a similar scenario for the head-initial tendency could have stemmed from neighbouring languages to the south of Chinese, namely Tai-Kadai, Hmong-Mien and Austroasiatic, which consistently use head-initial structures ([Bennett 1979](#); [de Sousa 2015](#); [Szeto and Yurayong 2022](#)). It remains disputable whether it ultimately was the northern or southern part of the Chinese dialect continuum which has changed from the Proto-Chinese word order patterns. Furthermore, the investigation of historical syntax of Chinese with evidence solely from Archaic Chinese may also be biased towards several specific varieties of Chinese (as will be discussed in Section 2.3).

Despite uncertainty in the directionality of changes in Chinese, what has been agreed upon among Chinese contact linguistics scholars is that the role of language contact is certainly considerable with regards to both the north and the south. Based on the contact explanation, we can draw the line along the Qinling Mountain–Huaihe River Line (see [Szeto and Yurayong 2021](#)), a boundary which is also frequently used in other biosciences and population sciences to distinguish the two major vegetation, climate and demographic zones of China (see e.g., [Fang et al. 2002](#); [Xie et al. 2004](#)), as illustrated in Figure 2.



**Figure 2.** The Qinling Mountain–Huaihe River line (Szeto and Yurayong 2021, p. 557).

Furthermore, a more fine-grained classification of Chinese languages in terms of using a large set of word order features also identifies the transitional zone in between (Norman 1988; Chappell 2015; Szeto and Yurayong 2021), in which the preferred head-final tendency in the north and the head-initial one in the south are used interchangeably. The validity of the Qinling Mountain–Huaihe River Line will be discussed further in Section 5.4.

### 2.3. Archaic Chinese Word Order

Moving from the typological to philological approach, studies on historical Chinese syntax are, unlike the majority of the world's languages, possible thanks to somewhat sufficient documentation of early forms of Chinese, providing concrete diachronic data of word order structures. Despite a potential bias toward the dialects or sociolects of the text producers, the attested documents are useful in the sense that they confirm the non-harmonious word order profile of Chinese to be in place since the late-2nd millennium BCE. In this section, we deal with Archaic Chinese, a Chinese language form spoken from the Shang dynasty during the 14th–11th centuries BCE to the Warring States period during the 5th–2nd centuries BCE (see also the complete periodisations in Sun 1996; Chappell 2001; Aldridge 2013).

As discussed in Section 2.2, Chinese languages consistently show a head-final tendency in noun phrase structures where possessors and modifiers precede head nouns, a pattern which largely applies also to Archaic Chinese (Aldridge 2015a), as in (2).

- (2) 先                      王                      之                      道  
      former                king                    that[LINK]        way  
      ‘ways of the former kings’ (Analects, Xue'er, Aldridge 2013, p. 44)

As in modern Chinese languages, modifiers are often attached to head nouns with a linker, a demonstrative 之 ‘that’ in the Archaic Chinese case. Likewise, relative clauses precede the head noun, and the two constituents are linked by the demonstrative 之, as in (3).

- (3) 避                      世                      之                      士  
      avoid                world                that[LINK]        scholar  
      ‘scholar who avoids the world’ (Lúnyǔ Wēizǐ, Aldridge 2015a, 3b)

When there are multiple modifiers in a noun phrase, there is at least evidence for demonstratives preceding quantifiers, as in (4).

- (4) 願 君 去 此 三 子 者 也。  
 desire lord dismiss [this 3 gentleman] DET NMLZ  
 ‘(I) hope your lordship will dismiss these three gentlemen.’ (*Hánfēizǐ* 36, Aldridge 2015a, 4b)

Interestingly, the use of classifiers was not attested in the Archaic Chinese sources (Peyraube 1991), but the order was consistently head nouns in the final position, as ‘these three gentlemen’ in (4).

The use of prepositions, meanwhile, has been attested since the Archaic Chinese period when locational and dynamic verbs were already grammaticalised and used in the preverbal position, such as 於 ‘(in)to’ (< ‘to go’), 在 ‘in, at’ (< ‘to be in’) and 從 ‘from’ (< ‘to follow’), as can be seen in (5).

- (5) 孝文 帝 從 代 來。  
 Xiàowén emperor [from Dài] come  
 ‘Emperor Xiàowén arrived from Dài.’ (*Shǐjì Xiàowén běnjì*, Aldridge 2015a, 6a)

Typologically speaking, prepositions should be more common among head-initial languages and can be considered a secondary innovation in head-final Chinese languages, supplementing a more typical postpositional phrase based on relational nouns, such as 上 ‘top’ and 面 ‘side’, a strategy which is still productive in modern Chinese languages and other Sino-Tibetan languages (see concrete examples in Section 4.3).

In terms of the verb phrase, verb-medial basic word order has been attested since the beginning of literary Chinese language history. However, pragmatic variation due to dislocation of topic has also been attested since the Archaic Chinese period. Often, the topicalised lexical unit may be repeated with a resumptive pronoun, such as the demonstrative-derived 3rd person pronoun 之 in (6).

- (6) 子路, 人 告 之 以 有 過。  
 [Zǐlù]<sub>TOP</sub> person tell [3.ACC]<sub>RES</sub> APPL have error  
 ‘Zǐlù, someone told him he made a mistake.’ (*Mèngzǐ Gōngsūnchǒu shàng*, Aldridge 2015a, 7)

In any case, there are also instances of the verb-final pattern in some specific constructions, which correspond to equivalent constructions in other Sino-Tibetan languages and are thus viewed by several scholars as remnants of pre-Archaic Chinese syntax (e.g., Li and Thompson 1974; Feng 1996; Xu 2006). This concerns, for instance, non-in situ question word phrases (7) and negative clauses (8). As for objects in negative clauses, there is a diachronic observation that fronting of the object disappeared later towards the 1st century BCE, as contrasted between texts from the 5th century BCE (8a) and the 1st century BCE (8b) (Aldridge 2015a).

- (7) 公 誰 欲 與 \_\_\_\_?  
 2<sub>SG</sub> who want give  
 ‘Who do you want to give (it) to?’ (*Zhuāngzǐ Xúwúguǐ*, Aldridge 2015a, 9a)
- (8) a. 莫 我 知 也 夫!  
 none [1<sub>SG</sub> know] NMLZ EXCL  
 ‘No one understands me!’ (*Lúnyǔ Xiànwèn*, the 5th century BCE, Aldridge 2015a, 18a)
- b. 莫 知 我 夫!  
 none [know 1<sub>SG</sub>] EXCL  
 ‘No one understands me!’ (*Shǐjì Kǒngzǐ shìjiā*, the 1st century BCE, Aldridge 2015a, 18b)

Some previous studies have argued that the reason for object fronting may be due to an information structure effect (Djamouri 1991), cliticisation of the object (Feng 1996) or a need to apply accusative case marking on the object (Aldridge 2015b). Given the lack of



consensus on the factor(s) behind dislocation of the object, we maintain in the current study that pragmatic factors are significant for determining variation but are at the same time difficult to predict and judge from the ancient texts with the limited contexts given. In any case, we still consider it useful to discuss the order of direct object and verb in comparison with other Sino-Tibetan languages with a rigid verb-final pattern (see also an additional explanation in Section 3).

Regarding other verb phrase structures, there is a consistent division between head-final and head-initial patterns. Head-initial modal auxiliaries (9) and complement clauses (10) consistently occur since the Archaic Chinese stage. Meanwhile, locational adverbials and temporal adverbials may vary between the preverbal (5) and postverbal positions (9), with both being equally frequent (Aldridge 2015a). Preverbal adverbials likely stem from the grammaticalisation of locational and dynamic verbs, as discussed above.

- (9) 天子            能            薦            人            於            天。  
ruler            [can            recommend] person            [to            heaven]  
'The ruler can recommend someone to heaven.' (*Mèngzǐ Wānzāng shàng*, Aldridge 2015a, 5a)
- (10) 言            非            禮            義,            謂            之            自            暴            也。  
speech betray Rite Righteousness say [3.ACC self injure NMLZ]  
'If his speech betrays the Rites and Righteousness, then (one) says of him that he harms himself.' (*Mèngzǐ Lǐlǒushàng*, Aldridge 2015a, 22)

As for comparatives, the head-initial construction with adjectives preceding standards of comparison linked by comparative markers 於 'to go (into)' (11a) or 過 'to surpass' (11b) has been attested in Archaic Chinese (Sun 1996, pp. 10–11).

- (11) a. 苛            政            猛            於            虎。  
severe government [ferocious COMP tiger]  
'A severe government is more ferocious than a tiger.' (*Lǐjì*, Sun 1996, p. 25)
- b. 由            也            好            勇            過            我。  
Yóu PART fond [dare COMP 1sg]  
'Yóu is more fond of daring than I (am).' (*Lúnyǔ Gōngyě zhǎng*, Sun 1996, p. 38)

Later, another head-final construction with standards of comparison preceding adjectives emerged in Middle Chinese (the 2nd century CE onwards), through functional extension from a lexical meaning of 比 'to compare' in Archaic Chinese (12) to a comparative marker in Middle Chinese (Sun 1996, p. 39).

- (12) 爾            何            曾            比            予            於            是?  
2sg how EMPH compare 1sg to 3sg  
'How (dare) you compare me to him?' (*Mèngzǐ Gōngsūn Chōu Shàng*, Sun 1996, p. 39)

The Middle Chinese alternation is still productive in southern Chinese languages, while the head-final construction with 比 'to compare' is the only pattern in northern Chinese languages (Sun 1996, p. 38; Ansaldo 1999, 2010; Chappell 2015). Interestingly, the head-final type is used across other Sino-Tibetan languages (see Section 4.3), so the Archaic Chinese head-initial construction with 於 'to go (into)' and 過 'to surpass' showed an early deviation from the head-final baseline tendency. At the same time, the preferred patterns in the present-day northern and southern Chinese languages align with those of their neighbouring Altaic languages in the north and Mainland Southeast Asian languages in the south, as discussed in Section 2.2.

Through the discussion with evidence from Archaic Chinese in this section, we see that a non-harmonious profile of Chinese with hybrid word order is diachronically consistent. This observation has motivated the idea that Chinese was originally a syntactically hybrid language which likely has maintained the Sino-Tibetan head-final tendency to a certain extent especially in noun phrase structures but also acquired head-initial structures from neighbouring languages. The head-initial tendency may be subject to areal diffusion from the south, particularly Tai-Kadai and Hmong-Mien, which have been in contact with Chinese for several millennia and from which a significant number of groups have shifted

their language to Chinese (see also [DeLancey 2013](#) for the discussion on the mixed origin of Chinese). Though such sociolinguistic factors and demographic changes may also be responsible for the non-harmonious word order patterns in Chinese throughout its attested history, we speculate that extralinguistic evidence and related explanation may also contribute to the discussion of the word order typology of Chinese and its change from a broader perspective.

#### 2.4. Sociolinguistic and Other Explanation to Language Change

In addition to linguistic reconstruction methods, we also consider the role of sociolinguistic factors in language variation and change. Social factors alone cannot determine variation in language structure, and they are often difficult to separate from inheritance. Nevertheless, several theoretical approaches have looked for a link between sociolinguistics and variation in structural features of languages (as discussed in [Ladd et al. 2015](#)). One of them is sociolinguistic typology, a research program initiated by [Trudgill \(2011, 2020\)](#). Sociolinguistic typology aims at bridging the gap between sociolinguistics and linguistic typology, which up until recently have been largely separate subfields of linguistics. Sociolinguistics has mainly focused on language-external factors that shape language use and cause variation (such as age, social class, gender, language policies), while linguistic typology has been more concerned about language-internal diversity (such as word order correlations). A key question in sociolinguistic typology is whether there are any systematicities regarding what types of linguistic structures are developed and maintained in different types of sociolinguistic environments. Sociolinguistic typology has looked at, for example, the role of community size (small vs. large), density of social networks (dense vs. loose), social stability (stable vs. instable), the degree of shared information in the language community (high vs. low) and degree of contact with neighbouring language communities (high degree of contact vs. isolation) in relation to the complexity of language structure.

Recent research on sociolinguistic typology has shown that population size, the number of native speakers (henceforth, L1 speakers) and the proportion of second language speakers (henceforth, L2 speakers) correlate with morphological complexity. By using data from *World Atlas of Language Structures* (WALS), [Lupyan and Dale \(2010\)](#) have shown that languages with a smaller number of native speakers tend to have more complex morphological systems than languages with a larger number of native speakers. [Bentz and Winter \(2013\)](#) have found evidence that the greater the proportion of L2 speakers in the speech community is, the smaller the case system of the language is. [Sinnemäki and Di Garbo \(2018\)](#) have argued that morphological complexity depends on both population size and the number of L2 speakers. Finally, [Sinnemäki \(2020\)](#) has studied the role of population size and the number of L2 speakers together with a language-internal factor, word-order, in predicting the number of cases in a language. His results suggest that both population size and the number of L2 speakers are important predictors of the number of cases in a language, either in complex interaction with word order or as an independent predictor together with word order ([Sinnemäki 2020](#)).

Research on sociolinguistic typology has also found some evidence that a large population size and a high number of adult L2 speakers can lead to the preference of analytic expressions instead of inflectional expressions. A case in point is the reflexive in Kuki-Chin and Bodo-Garo languages of Northeast India ([DeLancey 2014](#)). Proto-Tibeto-Burman originally had a middle reflexive suffix *\*(n)si*, which has been lost in both Kuki-Chin and Bodo-Garo subbranches. However, both subbranches have recreated new reflexive constructions, whose structures at least partially reflect the sociolinguistic contexts in which these languages are spoken. Kuki-Chin languages, spoken by small populations in isolated hill locations, have created an inflectional middle reflexive suffix *-a*, e.g., *ka-a-thooŋ* [1sg-REFL-hit] ‘I hit myself’. Contrary to Kuki-Chin, Bodo-Garo languages are spoken by large populations in the Assam Valley in Northeast India. They serve as lingua francas in the region and are therefore influenced by contact from large numbers of L2 speakers who are learning them as adults. Bodo-Garo languages use an analytic expression, an emphatic

pronoun *gau*, to express reflexivity, e.g., *bi-w gau-khwu aina-ao nai-dwŋ-mwn* [3sg-SUBJ self-OBJ mirror-LOC see-REAL-PST] ‘He saw himself in the mirror’. One of the possible reasons why Kuki-Chin and Bodo-Garo languages are using different strategies for reflexive expression is that analytic expression is easier to learn for adult L2 speakers. Therefore, sociolinguistic factors such as population size and multilingualism shape the environment in which the language is learned and used, and this, in turn, can affect the language structure together with internal factors.

Apart from bilingualism, language attitudes and ideologies, language awareness and stylistic practice have been shown to be important in understanding language change (Rodríguez-Ordóñez 2019). While typologically similar languages tend to converge in language contact and typological distance tends to deter convergence (see Thomason and Kaufman 1988), in some contact situations the language attitudes of the speakers may override the effects of structural similarities or differences in the languages in question. In any case, research on attitudes and ideologies has been largely qualitative and difficult to quantify, so we do not include it as a variable in this study.

As the language sociology of a given community not only concerns its members but also its environment, we must not ignore the role of geographical factors such as climate and mobility. A pioneer work by Johanna Nichols (1992) and subsequent publications with various collaborators have identified typological clines which are the result of early human expansion prior to the formation of current linguistic groups or families and *Sprachbünde*. Most notably, the observation points to a stable incline or decline within selected language features from the west to the east of the northern hemisphere, especially across Northern Eurasia. However, such a large-scale approach has not been applied much in the quantitative explanation of linguistic diversity and variation of Sino-Tibetan languages spoken across the Trans-Himalayan zone. The current study operates on a working hypothesis developed by Szeto and Yurayong (2021) that Chinese speaking areas potentially illustrate the incline or decline of typological features in a stable and statistically significant manner. By integrating word order and geographical data, we seek to identify another north-to-south and east-to-west cline in East Asia and show that variation across Chinese, as discussed in previous studies and given in Section 2.2, are not necessarily divided sharply by the Qinling Mountain–Huaihe River line, but the decline of the head-final tendency has a significant correlation with extralinguistic factors (see Section 5.4).

We have seen in this section that both extralinguistic and intralinguistic factors are important in understanding language variation and change. The extralinguistic factors which have received most attention in studies of sociolinguistic typology are speaker population in terms of size and proportion of L1 and L2 speakers, as well as geographical factors. These factors have been shown to play an important role in predicting morphological complexity and patterns of language change. However, sociolinguistic factors alone cannot predict language change, and they must be studied together with diachronic analysis and language-internal factors, such as word order correlations discussed in Sections 2.1–2.3.

### 3. Data, Method and Results

The data for comparative analyses in the present study include 20 Chinese and 43 Tibeto-Burman languages, acquired from published reference grammars, grammatical sketches and analyses of individual constructions. The sample includes one language of each Sino-Tibetan second level subbranch as classified in Glottolog 4.6 (Hammarström et al. 2022). Several methods have been practised in linguistic typology with different degrees of control for genealogical and areal biases. For the present study, we adapt the variety sampling method developed by Miestamo et al. (2016), which takes one language per one branch. The reason for choosing this method is to capture general tendencies across Sino-Tibetan and to prevent an issue of oversampling, in which Chinese with the highest numbers of speakers and varieties would become unnecessarily overrepresented for the entire Sino-Tibetan family (see also discussion on advantages and issues within different sampling methods in Guzmán Naranjo and Becker 2022).

Since the milestone of the present study is to identify if there exist statistical correlations between the degree of directionality of change in word order and sociolinguistic factors, it is also crucial that sociolinguistic information such as numbers of speakers and geographical locations of linguistic varieties and communities under investigation are available for data accumulation. Information on genealogical status of individual variety in the Sino-Tibetan language family, number of speakers, geographical location in GIS coordinators and sources of data is given in Supplementary Material A. However, in reality, the data on word order features and sociolinguistic information may be insufficiently or fragmentally provided in the grammatical descriptions for some subbranches. This is the case for Gongduk, which unfortunately is excluded from the sample in terms of the quantitative analysis. In any case, Gongduk examples will still be mentioned in the qualitative analysis when they are relevant for the discussion.

In terms of word order features, the current study investigates 16 constructions involving different types of phrases and clauses, as given in Table 3.

**Table 3.** Word order features under investigation.

Feature	Construction	Head-Dependent	Domain
1	Noun compounding	Noun[head]–Noun[modifier]	Noun (Section 4.1)
2	Adjectival modification	Noun–Adjective	
3	Adnominal possession	Possessee–Possessor	
4	Gender specification	Noun–Gender specifier	
5	Quantity modification	Noun–Numeral	
6	Deictic modification	Noun–Demonstrative	
7	Noun relativisation	Noun–Relative clause	Adjective (Section 4.2)
8	Degree adverb	Adjective–Degree adverb	
9	Comparative	Adjective–Standard of comparison	
10	Locational adverbial	Verb–Locational adverbial	Verb (Section 4.3)
11	Direct object	Verb–Direct object	
12	Predicative possession	Verb–Possessee	
13	Modal auxiliation	Verb[modal]–Verb[content]	
14	Adposition	Adposition–Noun	
15	Reported speech	Speech verb–Complement clause	
16	Negation	Negator–Verb	

The analysis takes into account the relation between the head and dependent of the constructions, thereby enabling a quantification of head-initial vs. head-final tendencies occurring in a given language.

It is worth remarking that direct objects can be prone to variation due to pragmatic factors, such as information structure (topic and comment) and text genres (narrative and dialogue), as discussed in Section 2.3 in connection to (6) and (8). A similar issue of pragmatically variable syntactic constructions, which causes some difficulties in terms of comparison and syntactic reconstruction, has been also discussed for Austroasiatic languages (see Jenny 2020). Despite such variation, it is still reasonable to consider this as a feature which is expected to appear as a head-initial construction in languages with more loosened tendency of verb-final structures.

In the process of data collection, queries are biased toward the head-final variant of a construction, and data are organised as binary values: 1 = head-final vs. 0 = head-initial. In cases where variation is reported in language descriptions, i.e., a language allows both head-final and head-initial variants of a construction either conditionally in specific contexts or unconditionally, the value is assigned as 0.5. These word order scores aggregated



from different constructions under investigation have two uses. On the one hand, they serve as evidence for the generalisation of cross-family tendencies and a baseline for the reconstruction of Proto-Chinese and Proto-Sino-Tibetan syntax (see Section 5.3). On the other hand, they are used to identify whether there exist statistical correlations between the degree of word order variation and extralinguistic factors related to language sociology of given speech communities (see Section 5.4).

Supplementary Material B shows the results with head-finalness scores of each variety in the right-most column and family-average scores of each word order features in the bottom lines. As Chinese languages make up more than one-third of the sample, the feature average scores are accumulated separately from the rest of the Tibeto-Burman languages to prevent bias toward Chinese.

Several immediate observations can be made from Supplementary Material B, which will be followed up in more details in Section 4. Among Chinese, the range of variation is between 9.5 and 13.5, with Zhoutun being on the top. More strikingly, Karenic languages have significantly low overall head-finalness scores, ranging from 4.5 to 6.5. This corresponds to previous observations regarding a radical typological shift in Karenic which could have taken place already at the Proto-Karenic stage based on the observed distribution (see Kato 2019 and further observation and discussion in Sections 4 and 5, respectively). At the same time, overall head-finalness scores are consistently high across Bodic, Dhimalish, Himalayish, Macro-Tani and Raji-Raute languages. Feature-wise, adnominal possession is consistently head-final across languages, followed by noun compounding, locational adverbial, relative clause and comparative constructions, in line with earlier typological investigations of variation across Chinese languages (Chappell et al. 2007, p. 189; Szeto and Yurayong 2021).

A closer look will be taken at individual word order features with concrete examples in Section 4, focusing on how they are generally realised across Chinese and Tibeto-Burman languages and whether some specific branches show significant deviation from the baseline.

#### 4. Description of Word Order Variation across Sino-Tibetan Languages

This section provides examples of 16 word order features under investigation by primarily describing homogeneity or variation across Chinese languages and secondarily making references to Tibeto-Burman languages as a supplementary explanation of individual features. Table 4 contrasts the average head-finalness scores between Chinese and other Sino-Tibetan languages to support the description below in this section.

The results are largely in line with observations in the previous cross-Sino-Tibetan surveys by Dryer (2003) and LaPolla (2015) in that the head-initial tendency is present in several grammatical constructions, deviating from the head-final baseline. In particular, adjectival and quantity modification show a consistent opposition between Chinese, which prefers the head-final pattern, and Tibeto-Burman, which prefers the head-initial pattern, while the opposite applies for predicative possession, modal auxiliaries and reported speech. Obvious and subtle differences between Chinese and other Sino-Tibetan languages will be discussed below in connection with individual features. Unless the source is given for language examples embedded in text, particularly the Chinese ones, refer to the sources given in Supplementary Material A.

**Table 4.** Average head-finalness scores for Chinese and Tibeto-Burman languages: 1 = head-final vs. 0 = head-initial.

Feature	Construction	Chinese	Tibeto-Burman	Domain
1	Noun compounding	1.00	0.95	Noun (Section 4.1)
2	Adjectival modification	0.98	0.36	
3	Adnominal possession	1.00	0.98	
4	Gender specification	0.35	0.06	
5	Quantity modification	0.93	0.20	
6	Deictic modification	0.98	0.71	
7	Noun relativisation	1.00	0.77	Adjective (Section 4.2)
8	Degree adverb	0.83	0.65	
9	Comparative	0.83	0.91	
10	Locational adverbial	1.00	0.86	Verb (Section 4.3)
11	Direct object	0.53	0.88	
12	Predicative possession	0.10	0.93	
13	Modal auxiliaries	0.43	0.94	
14	Adposition	0.55	0.94	
15	Reported speech	0.08	0.71	
16	Negation	0.05	0.33	

#### 4.1. Noun Phrase Structures

Across Chinese and Tibeto-Burman languages, the most consistently head-final noun phrases are compounds and possession. In Chinese noun compounds, a modifier noun always precedes a head noun, e.g., Shanghai Wu *jiéu dí* [wine **shop**] ‘hotel’, *shíā sǐ* [fragrant **water**] ‘perfume’, and *hú tsuo* [fire **car**] ‘train’ (see also Arcodia 2007 for further subtypes of compounds). This is also a general tendency across Tibeto-Burman languages, but variation is, however, marginally found, e.g., Sgaw Karen *pli t<sup>h</sup>ā<sup>7</sup>* [string **iron**] ‘wire’ vs. *p<sup>h</sup>ē t<sup>h</sup>ū* [necklace gold] ‘golden necklace’ (Kerbs Forthcoming), and Anong *k<sup>h</sup>en<sup>55</sup> ts<sup>h</sup>h<sup>31</sup>* [vegetable **juice**] ‘vegetable soup’ vs. *luŋ<sup>55</sup> su<sup>55</sup>* [stone mill] ‘grindstone’ (Sun and Liu 2009, p. 50).

Likewise, a head-final adnominal possessive construction with the possessor preceding the possessee is prevalent across Chinese and Tibeto-Burman, e.g., Kunming Mandarin *ni<sup>3</sup> nā<sup>1</sup> tie<sup>1</sup>* [2SG LINK **father**] ‘your father’ and *gio<sup>2</sup> cío<sup>4</sup> nā<sup>1</sup> ts<sup>h</sup>iē<sup>2</sup> ts<sup>h</sup>ā<sup>3</sup>* [school LINK **property**] ‘school’s property’. Juxtaposition is a common strategy, while some languages may also optionally use genitive linkers, e.g., Mandarin *de* 的 and Cantonese *ge<sup>3</sup>* 嘅, which can be an instance of alienable possession in some sense (see e.g., Li 2018, pp. 54–57 for Yichun Gan). However, there is also an instance of head-initial adnominal possession reported in Raji with possessors marked by possessive suffixes, e.g., *tsa-ŋ* [son-POSS.1SG] ‘my son’, and this pattern has likely been adopted from contact with an Indo-Aryan language (Khatrī 2008, p. 22).

In contrast, there are several noun modifiers which are to a certain extent consistently placed before a head noun in Chinese but after a head noun in Tibeto-Burman, such as adjectives, gender specifiers and quantifiers. While prenominal adjectives and quantifiers are prevalent across Chinese languages, e.g., Yichun Gan *san<sup>34</sup> pun<sup>42</sup> xao<sup>42</sup> cy<sup>34</sup>* [three CLF good **book**] ‘three good books’, northwestern Mandarin also allows postnominal adjectives as in Wutun and postnominal quantifiers as in Tangwang, Wutun and Zhoutun, e.g., Wutun *hu yak-la~la-de-ge* [flower beautiful-INCOMP~INCOMPL-NMLZ-CLF] ‘beautiful flower’ and *gek san-ge* [dog three-CLF] ‘three dogs’. This phenomenon has been explained for these varieties as a result of contact with Tibetan. For Tibeto-Burman, a diachronic scenario is the reanalysis

of relative clauses (see Section 5.1). Interestingly, some other factors for variation are also reported in language descriptions or noticed in our observation.

In terms of semantics, when there are multiple adjectives in a noun phrase in Konyak, a language which allows the use of prenominal adjectives (prefixed by *a-*) alongside canonical postnominal adjectives (suffixed by *-pu*), there is a tendency to place a quality adjective before the head noun but a quantity adjective after noun, e.g., *yəwməy-pu ciŋ a-n̄u* [beautiful-ADJ village ADJ-big] ‘a beautiful big city’ (Nagaraja 2010, pp. 75–76). From the syntactic perspective, meanwhile, East-Central Tangkhul Naga, a language which also allows free alternation between prenominal and postnominal adjectives, seems to prefer the prenominal adjectives in the contexts of predicative (13) and reduplication for plurals e.g., *k<sup>h</sup>ra k<sup>h</sup>ra seiŋ* [old old house] ‘old houses’, while the word order for adjectives in subject and direct object phrases remains more variable (see Devi 2011, pp. 134–40, 231, 234, 290–91).

#### East-Central Tangkhul Naga (Kuki-Chin-Naga)

- (13) a. *və* *kəp<sup>h</sup>ə* *ləsɪnəu* *a-ŋi-mə-ne.*  
 3SG.FEM [good girl] NEG-BE-NEG-ASP  
 ‘She is not a good girl.’ (Devi 2011, p. 256)

Regarding phonology, Anong, which predominantly uses postnominal adjectives, may also allow the use of adjectives with two or more syllables in the prenominal position before a nominaliser *u<sup>55</sup>*, e.g., *ba<sup>35</sup>ba<sup>31</sup>-tɕ<sup>h</sup>ɛn<sup>31</sup> u<sup>55</sup> ʂ<sup>55</sup>va<sup>31</sup>* [thin-DIM NMLZ book] ‘a thin book’ and *sɪ<sup>31</sup>la<sup>33</sup> u<sup>55</sup> a<sup>31</sup>ts<sup>h</sup>an<sup>31</sup>* [good NMLZ person] ‘a good person’ (see Sun and Liu 2009, p. 115).

Specification of gender shows two patterns across Chinese languages, among which the northern varieties have a head-final pattern (14a), whereas the southern varieties have a head-initial pattern (14c). Interestingly, Changsha Xiang allows both orders as free alternation among different animal referents, first attested in *Shimén Xiànzhi* ‘Gazetteer of Shimen’ in 1875 (Wu 2005, p. 113). This may be largely due to the transitional identity of Xiang and its geography in the central zone of the Chinese dialect continuum (as discussed by Ho 1987; Norman 1988, p. 182; Szeto and Yurayong 2021, and in Section 2.2).

- (14) a. *gōng* *jī* *mǔ* *jī* Beijing Mandarin  
*h’iōng* *jī* *tsǐ* *jī* Shanghai Wu  
*kən<sup>33</sup>* *tɕi<sup>33</sup>* *po<sup>13</sup>* *tɕi<sup>33</sup>* Changsha Xiang  
*(s)kjo-* *kukəri* *mant-* *kukəri* Kinnauri  
 male chicken female chicken  
 ‘rooster’ ‘hen’
- b. *daŋkha* *juhā* *maini* *juhā* Dhimal  
 male rat female rat  
 ‘male rat’ ‘female rat’
- c. *tɕie<sup>11</sup>* *kan<sup>11</sup>* *tɕie<sup>11</sup>* *ṃ<sup>24</sup>* Tunxi Hui  
*tɕi<sup>33</sup>* *kən<sup>33</sup> tsɪ* *tɕi<sup>33</sup>* *po<sup>13</sup> tsɪ* Changsha Xiang  
*ki<sup>21</sup>* *bo<sup>33</sup>* *ki<sup>21</sup>* *tɕie<sup>55</sup>* Caijia  
*naga* *-whaba* *naga* *-mama* Eastern Tamang  
 chicken male chicken female  
 ‘rooster’ ‘hen’

The head-final pattern is considered a Chinese construction, though our data show that it is consistently so only for Mandarin and Jin, while the other Chinese languages allow and use the head-initial pattern more frequently. The head-initial pattern is often considered to be a pattern borrowing from Mainland Southeast Asian substrate languages, most notably Tai-Kadai and Hmong-Mien (Yuan 1983, p. 10; Wang 1991, pp. 177–78; Pan 1991, pp. 287–88; Szeto and Yurayong 2021, p. 566). Such a preference is concretely reported for the Xiang dialect continuum, in which the northern part prefers the head-final pattern, which is considered to be prototypical for Chinese, while the southern part prefers the

head-initial pattern, considered an innovation (Wu 2005, pp. 111–13). At the same time, Tibeto-Burman languages mostly use the head-initial pattern like in southern Chinese languages (Szeto and Yurayong 2022, p. 29), with the exception of the head-final pattern observed in Dhimalish languages and Kinnauri, as shown in the parallel examples in (14a) and (14b). The emergence of the head-initial structure in both Chinese and Tibeto-Burman can also be explained language-internally without unnecessary reference to a contact explanation, but rather to the nominalisation of the adjectival constituent (see further discussion in Section 5.1). As a side note, we also observe a conditioned alternation in Toto, as given in (15).

#### Toto (Dhimalish)

- |         |  |   |
|---------|--|---|
| (15) a. | <i>dabe-kunwa</i><br>MASC-tiger<br>'tiger'                                   | <i>cabe-kunwa</i><br>FEM-tiger<br>'tigress'   |
| b.      | <i>bale-keka</i><br>MASC-chicken<br>'cock'                                   | <i>cabe-keka</i><br>FEM-chicken<br>'hen'  |
| c.      | <i>muri</i><br>chilli<br>'male chilli' (a chilli which fails to bear fruits) | <i>dabe</i><br>male   |
| d.      | <i>yui-wa-poja</i><br>dance-AGENT-MASC<br>'male dancer'                      | <i>yui-wa-meme</i><br>dance-AGENT-FEM<br>'female dancer' (Basumatary 2016, pp. 81–83) |

The rule is that Toto uses gender specifiers before animal (15a) and bird referents (15b) but after plant (15c) and human referents (15d).

Regarding the use of numerals and classifiers, some conditioned behaviour is observed in several languages which allow both prenominal and postnominal quantifiers. For instance, Eastern Tamang classifiers are obligatory when the quantifier precedes a noun (16b), but they can be omitted when the quantifier follows a noun (16a), the latter being reportedly a more frequent pattern in texts and normal dialogue (Lee 2011, p. 32).

#### Eastern Tamang (Bodic)

- |         |   |                            |                               |                          |
|---------|---|----------------------------|-------------------------------|--------------------------|
| (16) a. | <i>jha-gade</i><br>son-PL                           | (gor)<br>[(CLF)            | <i>som</i><br><b>three</b>    | <i>mu-la</i><br>COP-NPST |
|         | '(He) has three sons.' [lit. There are three sons.] |                            |                               |                          |
| b.      | <i>gor</i><br>[CLF                                  | <i>som</i><br><b>three</b> | <i>jha(-gade)</i><br>son(-PL) |                          |
|         | 'three sons' (Lee 2011, p. 32)                      |                            |                               |                          |

At the same time, Amri Karbi generally uses fused forms of numerals attached to classifiers in the postnominal position, e.g., *kampi i-jon* [monkey **one-CLF**] 'one monkey', but also sometimes allows the use of prenominal quantifiers for animate referents, e.g., *isi i-jon a-kampi-so* [**one one-CLF** POSS-monkey-DIM] 'one little monkey' (Philippova 2021, pp. 129–30). The alternation is also reported for Garo *mechik sak-sa* [woman **CLF-one**] and *sak-sa mechik* [**CLF-one** woman] 'one woman' (Burling 2003, p. 97), and Zakhriung *dungpu nga* [tree **five**] 'five trees' and *nga simjong* [**five** banana] 'five bananas' (Landi 2005, p. 56).

Other groups of noun modifiers with less consistent variation between head-final and head-initial patterns are demonstratives and relative clauses. On the one hand, demonstratives are prenominal across Chinese languages, e.g., Xi'an Mandarin *tʂʰʅ<sup>55</sup>/u<sup>55</sup>/næ<sup>55</sup> kɿ<sup>31</sup> zən<sup>31</sup>* [this/that/yonder CLF person] '**this/that/yonder person**'. On the other hand, there is variation between prenominal and postnominal demonstratives among Tibeto-Burman languages. Interestingly, for noun phrases with multiple modifiers in Toto, demonstratives occur consistently in the prenominal position, while adjectives may alter between the prenominal (17a) and postnominal positions (17b).



**Toto (Dhimalish)**

- (17) a. *i*                *dasiwa*                *ziya*                *u*                *haŋpuwa*                *ziya*  
           **this**                **black**                bird                **that**                **white**                bird  
       b. *i*                *ziya*                *dasiwa*                *u*                *ziya*                *haŋpuwa*  
           **this**                bird                **black**                **that**                bird                **white**  
           ‘this black bird’                                ‘that white bird’ (Basumatary 2016, pp. 152–53)

Such a phenomenon is not observed in Chinese languages because the position of adjectives is consistently prenominal (as discussed above).

As for relative clauses, their position is always prenominal in Chinese languages, as shown in (18a). Note that Tunxi Hui speakers reportedly use more spontaneously a construction formed by a noun phrase with a demonstrative without a relativiser as in (18b) (Lu 2018, pp. 167–68).

**Tunxi Hui**

- (18) a. *kə*<sup>44</sup>                *ɕio*<sup>11</sup>                *ka*                *tʂʰə*<sup>42</sup>  
           [3SG                cook                REL]                **dish**  
           ‘the dishes which (s)he cooks’  
       b. *tɕʰi*<sup>25</sup>                *tɕiɲ*<sup>24</sup>                *mo*<sup>31</sup>                *ka*<sup>42</sup>                *ian*<sup>44</sup>  
           [eat                dumpling]                **that**                CLF                **person**  
           ‘the person who is eating dumplings’ (Lu 2018, p. 169)

The predominance of prenominal relative clauses also applies to Tibeto-Burman. There is, however, an exception in Karenic languages which predominantly use postnominal relative clauses for both subject (19a) and object relativization (19b). At the same time, the use of the prenominal pattern for objects is also reported for Eastern Pwo Karen under certain conditions (20).

**Sgaw Karen (Karenic)**

- (19) a. *pyākəŋs*    *lá*                *ʔə*                *hé*                *lá*                *pyākəŋs*    *kə*  
           **person**    [REL                3SG                come                LOC                Karen                country]  
           ‘the person who came to Kayin State’  
       b. *pyākəŋs*    *lá*                *jə*                *tə*                *ʔs*                *nê*  
           **person**    [REL                1SG                hit                3SG]                **that**  
           ‘that person whom I hit’ (Kato 2021, p. 356)

**Eastern Pwo Karen (Karenic)**

- (20) *jə*                *tháu*                *lá*                *dəu*                *phən*                *khánphài*    *nó*  
           [1SG                ride                LOC                room                inside]                **shoes**                **that**  
           ‘those shoes which I wear in the room’ (Kato 2021, p. 355)

Among other verb-final Tibeto-Burman languages, an unconditioned free alternation is also observed in several languages as indicated in Supplementary Material B. This is a consistent free alternation, for instance, in both Digarish languages as shown in the prenominal (21a) and postnominal patterns (21b).

**Tawra (Digarish)**

- (21) a. *hǎ*                *hibán*                *bóyà*                *jyinaŋdō*                *kitab*                *hənde*  
           I.NOM                [forest.DAT                go]                **cousin.DAT**                book.ACC                give.HAB.3SG  
           ‘I give the book to (my) cousin who goes to the forest.’  
       b. *masáŋsyígwèlàn*    *biríhiriso*                *cyá*                *katígharimso*  
           **tree.fruit.PL.ACC**    [fall.RECIP]                she.NOM                collect.RECIP  
           ‘She collected the fallen fruit.’ (Devi Prasada Sastry 1984, pp. 187–89)

**4.2. Adjective Phrase Structures**

Two constructions fall under adjective phrase: (1) degree adverb and (2) comparative. The use of an adverb ‘very’ varies across Chinese and Tibeto-Burman languages. Particularly, the northwestern Chinese languages predominantly use postadjectival degree adverbs, e.g., Zhoutun *ʃɿ li uɿ* [hot very] ‘very hot’. At the same time, some languages possess both preadjectival and postadjectival degree adverbs, as is the case for Dhimal preadjectival

particle *menəŋ* ‘very’ and suffix *-ŋ* ‘very’ (King 2008, p. 53), and Amri Karbi preadjectival particle *bohut* (potentially borrowed from Indic) and suffixes *-ad* and *-det* ‘very’ (Philippova 2021, pp. 74, 136, 192, 218).

Moving to the comparative, we are only interested in comparatives of degree, not in comparatives of (in)equality which may show more variation (see, e.g., Zhou 2023 for such variation in Zhoutun). The majority of Chinese languages can use a head-final pattern with a standard preceding an adjective, as in (22a). However, the far southern Chinese languages, Yue and Pinghua in particular, also allow the use of an adjective before a standard (23a), as contrasted and paralleled with other Tibeto-Burman languages in (22b) and (23b).

- |      |    |                         |                              |                  |               |
|------|----|-------------------------|------------------------------|------------------|---------------|
| (22) | a. | $a^{24}$                | $pi^{31}$ $k^h\partial^{44}$ | $ka^{11}$        | Tunxi Hui     |
|      | b. | $i$                     | $pa-sone$                    | $hubibe-lai$     | Bugun         |
|      |    | 1SG                     | COMP 3SG                     | 3SG-COMP         | tall(er)-DECL |
| (23) | a. | $\eta oi^{33}$          | $k\partial^{33}$             | $k\partial^{33}$ | $k'ui^{33}$   |
|      | b. | $j\bar{a}$              | $t^h\phi$                    | $-d\partial li$  | $s\bar{a}$    |
|      |    | 1SG                     | tall                         | COMP             | 3SG           |
|      |    | ‘I am taller than him.’ |                              |                  |               |

This contrast has been characterized as areal diffusion from neighbouring Altaic languages in the north and Mainland Southeast Asian languages in the south (Ansaldi 1999, 2010). The idea is also supported by the predominant head-final tendency across Tibeto-Burman languages, with the exception only for Karenic languages which use adjectives before standards of comparison, as is shown in (23b).

Typically across the Chinese languages, adjectives can precede a standard of comparison, even in Chinese languages which predominantly use the head-final *bǐ*-construction when the comparison involves degree measurement (Wu 2005, p. 183; Li 2018, p. 94; Lu 2018, p. 238), as shown in the Xiang example (24b).

#### Changsha Xiang

- |      |    |  |            |               |                                |
|------|----|--|------------|---------------|--------------------------------|
| (24) | a. | $t^h a^{33}$   | $pi^{41}$  | $\eta o^{41}$ | $kau^{33}$                     |
|      |    | 3SG  | COMP       | 1SG           | tall                           |
|      |    | ‘(S)he is taller than me.’                                     |            |               |                                |
|      | b. | $t^h a^{33}$   | $kau^{55}$ | $\eta o^{41}$ | $san^{33}$ $li^{13}$ $mi^{41}$ |
|      |    | 3SG  | tall       | 1SG           | [three centimetre]             |
|      |    | ‘(S)he is three centimetres taller than me.’ (Wu 2005, p. 183) |            |               |                                |

This pattern is considered by Yue-Hashimoto (2003, p. 111) as a reflex of the historical pattern, as discussed in Section 2.3 under (11). The study of Zhoutun comparatives also suggests that this pattern might be an instance of reinstalling the original head-initial comparatives (Zhou 2023).

#### 4.3. Verb Phrase Structures

Unlike nominal phrase structures, Chinese languages use many head-initial patterns whereas most Tibeto-Burman prefer head-final structures. For instance, most Chinese languages generally place the direct object after the verb in their basic word order, while also allowing the opposite pattern for topicalisation (as discussed in Sections 2.3 and 3). At the same time, the majority of Tibeto-Burman languages possess a verb-final basic word order, while several branches also allow placing the direct object after the verb, as is common for verb-medial Karenic and Caijia and less frequent in verb-final Zakhing and Mruic languages, as given in (25).

#### Mru (Mruic)

- |      |  |      |            |                 |
|------|--|------|------------|-----------------|
| (25) | $pari\eta$                                 | $ca$ | $wip^h um$ | $k^h\partial k$ |
|      | Paring                                     | eat  | mango      | PST             |
|      | ‘Paring eats mango.’ (Rashel 2009, p. 146) |      |            |                 |

The position for possessee in predicative possession also shows a direct correlation with basic word order. Chinese languages place the possessee in the postverbal position due to their verb-medial basic word order and *topic*-type of predicative possession, i.e.,

possessor as topic (see Stassen's 2009 typology of predicative possession). In any case, northwestern Chinese languages with the verb-final basic word order place the possessee before an existential verb, similarly to verb-final Tibeto-Burman languages. Interestingly the verb-medial Karenic languages can place the possessee before the existential verb, although the quantifier still follows the verb, as given in (26).

#### Geba Karen (Karenic)

- (26) *sā* *fī* *ʔ* *θó* *wà*  
 3.POSS house exist three CLF  
 'He has three houses.' [lit. His three houses exist.] (Naw 2008, p. 129)

This might be an instance of archaism as a remnant of verb-final pattern in Karenic (see further discussion in Section 5.3). In any case, the use of verb-medial pattern for predicative possession is also reported for Sgaw Karen when the construction belongs to the *with*-type in Stassen's (2009) classification (Kerbs Forthcoming), as given in (27a). This variant is considered by the informant as being more formal and poetic than the spoken variant with verb-final pattern given in (26) and (27b).

#### Sgaw Karen (Karenic)

- (27) a. *jə* *sē* *ʔóʔ* [more common,  
 1SG.POSS money exist spoken]  
 b. *jə* *ʔóʔ* *dʒʔ* *sē* [more formal,  
 1SG exist with money written, poetic]  
 'I have money.' (Kerbs Forthcoming)

As for adverbials, the expression of location is generally preverbal across Chinese and Tibeto-Burman languages, as is shown in (28).

#### Yichun Gan

- (28) *nɿ³⁴* *ʰshoe²¹³* *ko³⁴* *xau⁴²* *ko* *ʰai²¹³-²¹xoʔ* *ʰuʔcy³⁴* *a*.  
 2SG [LOC so good LINK university] study INTJ  
 'You are studying at such a good university!' (Li 2018, p. 72)

However, deviation is observed in Mruic languages which allow multiple clausal word orders, with the verb-medial pattern being more frequent (Rashel 2009, p. 160; Wright 2009, pp. 49–50), and crucially also alternation between preverbal (29a) and postverbal locational adverbials (29b). At the same time, Karenic languages consistently place locational adverbials after verbs (30) in accordance with their predominant verb-medial basic word order.

#### Hkongso (Mruic)

- (29) a. *pələŋkrum* *ʰaŋ* *hai* *aŋ* *toi* *ra*  
 [Paletwa side from] 1SG return come  
 'I come back from Paletwa.'  
 b. *va* *kəji* *ləmuk* *nam* *ʰaŋ*  
 bird fly [sky village at]  
 'The bird is flying in the sky.' (Wright 2009, p. 63)

#### Kayan Lahta (Karenic)

- (30) *kaɭjaŋʔ* *lwaŋ* *ni* *teŋ* *ba* *də* *tu*  
 Kayan go get porcupine CLF [in forest]  
 'The Kayan got a porcupine in the forest.' (Naw 2013, p. 79)

In any case, Chinese languages with verb-medial basic word order are homogenous in that they use both prepositions and postpositions. As discussed in Section 2.3, both are the results of grammaticalisation from locational or directional verbs to prepositions and relational nouns to postpositions, respectively, from the verb *lā⁵⁻⁴* 'to stay' and the noun *lai⁴* (<*lai⁴pak⁷*) 'inside', as shown in the Min example (31).

**Hui'an Min**

- (31) *kau*<sup>3</sup>                      *lā*<sup>5-4</sup>                      *tshu*<sup>5-3</sup>-*lai*<sup>4</sup>                      *khun*<sup>5</sup>  
 dog                      [at                      house-**inside**]                      sleep  
 'The dog sleeps in the house.' (modified from [Chen 2020](#), p. 262)

A similar phenomenon is also reported for Karenic, such as Eastern Pwo Karen *lā dāu phān* [LOC room **inside**] 'in the room' in (20), and Caijia *tu<sup>21</sup> ty<sup>24</sup> ty<sup>33</sup>* [at yard **upside**] 'in the yard' ([Lü 2022](#), p. 47). The observation from Chinese, Karenic and Caijia clearly points to a correlation between the verb-medial basic word order and prepositions (as discussed in [Greenberg 1963](#); [Dryer 1992a](#)). At the same time, the tendency that relational nouns always follow the content noun also correlates with the general head-final tendency of noun compounding (as discussed in Section 4.1). In the northwestern Chinese languages, Tangwang and Wutun, meanwhile, postpositions are the only option because verbs always occur in clause-final position and cannot precede any adpositional phrases. In any case, the verb-final Zhoutun also marginally uses prepositions, e.g., *iū xuṛthā* [along riverbank] 'along the riverbank' and *iū kṛ lu* [along this road] 'along this road' ([Zhou 2022](#), p. 106).

In terms of verb morphology, word order of the modal auxiliary is surveyed by looking at the position of an ability verb 'to be able'. Most Chinese languages head-initially place an ability verb such as 可以 before the main verb (32a), but some languages also head-finally use the postverbal auxiliary 得 'to get' for ability as in (32b).

**Cantonese**

- (32) a. *lei*<sup>5</sup>                      *ho*<sup>2</sup> *ji*<sup>5</sup>                      *daap*<sup>3</sup>                      *baa*<sup>31</sup>*si*<sup>2</sup>                      *heoi*<sup>3</sup>                      *man*<sup>4</sup> *faa*<sup>3</sup>                      *zung*<sup>1</sup>*sam*<sup>1</sup>  
 2sg                      [can                      catch]                      bus                      go                      culture                      centre  
 'You can take a bus (to get) to the Cultural Centre.'  
 b. *li*<sup>1</sup>                      *di*<sup>1</sup>                      *zi*<sup>1</sup> *liu*<sup>2</sup>                      *m*<sup>4</sup>                      *seon*<sup>3</sup>                      *dak*<sup>1</sup>                      *gwo*<sup>3</sup>  
 this                      PL                      information                      [NEG                      believe                      **get**]                      pass  
 'These figures are not worth believing.' ([Matthews and Yip 2011](#), pp. 265, 278)

[Paternicò and Arcodia \(2023\)](#) discuss variation between the use of 得 in preverbal and postverbal position and show that the postverbal use is predominant in Cantonese with more versatile modal functions than in Mandarin, in which the postverbal use of 得 is declining, whereas the preverbal use remains common.

At the same time, northwestern Chinese languages place the suffixed ability verb clause-finally as predicate, similarly to verb-final Tibeto-Burman languages, as given in (33) and (34).

**Zhoutun**

- (33) *tṣ*                      *ŋṣ*                      *tṣuṭhūxua*                      *itiātiā*                      *ṣuṣ=le=lɔ*  
 PART                      1sg                      Zhoutun.vernacular                      little                      [speak=able=PFV]  
 'I can speak a little Zhoutun vernacular.' ([Zhou 2022](#), p. 39)

**Bulu Puroik (Kho-Bwa)**

- (34) *atsi*                      *sāḍgo*                      *ap<sup>h</sup>ɔ*                      *ba-hi-rjaḍ-tfa*                      ...  
 grandchild                      Sandgo                      male                      [NEG-speak-able-PFV]  
 'Grandsons Sandgo's father doesn't know to speak [Puroik] ...' ([Lieberherr 2017](#), p. 188)

However, we find deviation in Mruic languages which place the ability verb before the main verb, as shown in (35). This pattern is similar to the use of preverbal 可以 'to be able' in Chinese languages, as in the Cantonese form *ho<sup>2</sup> ji<sup>5</sup>* in (32a).

**Hkongso (Mruic)**

- (35) *mu<sup>l</sup>mai<sup>l</sup>*                      *ma<sup>l</sup>*                      *nin<sup>l</sup>*                      *aŋ<sup>l</sup>*                      *kəcə<sup>l</sup>*                      *aŋ<sup>l</sup>*                      *no<sup>l</sup>*                      *hai<sup>l</sup>*                      *au<sup>l</sup>*  
 cloud                      SUBJ                      cover                      1sg                      if                      1sg                      [NEG                      **able**                      shine]  
 'If the clouds cover me, I cannot shine.' ([Wright 2009](#), p. 81)

Another feature related to verb morphology is negator, the order of which is predominantly preverbal in most Chinese languages as in (32b) and (36), as opposed to Tangwang and Zhoutun which also use postverbal negators in some contexts as in (37).



**Linxia Jin**

- (36) *tɕin11 zəʔ2 paʔ2 xa35 y42 lie11, t'ia11 tɕ'11 lie11-21!*  
 today [NEG fall] rain PART sky clear PART  
 'Today it does not rain, it is sunny!' (Wang 2007, p. 248)

**Zhoutun**

- (37) *tha i=kɔ ʃɕ sã=tɕi xuɔ=pu=xã a.*  
 1SG one=CLF person three=jin [drink=NEG=down] PART  
 'I cannot drink up three jin (of wine) alone.' (Zhou 2022, p. 75)

Although preverbal negators tend to be more common across Tibeto-Burman and the world's languages (see Dahl 1979; Dryer 2013), there are also languages which only use postverbal negators as indicated in Supplementary Material B, such as in (38).

**Northern Tujia (Burmo-Qiangic)**

- (38) *lai<sup>4</sup> nga<sup>2</sup> re<sup>2</sup> hu<sup>3</sup> ta<sup>1</sup>*  
 today 1SG wine [drink NEG]  
 'I will not drink any wine today.' (Braslett et al. 2006, p. 132)

At the same time, some Tibeto-Burman languages may also use circumflexive negators, being on both sides of the negated verb (39). This type of language is assigned with 0.5 in Supplementary Material B.

**Lepcha (Himalayish)**

- (39) *hó hryóp-pung ʔân mák-kung gang-lá taʔyu ʔáre ma-thop-ne*  
 2SG cry-PTCP and die-PTCP if-also girl this [NEG-get-NEG]  
 'Even if you cry or die, you won't get this girl.' (Plaisier 2006, p. 136)

As for negation, we also test whether negators and their order vary between main and subordinate clauses. This observation is a secondary finding which will be discussed in more detail with a diachronic analysis in Section 5.2.

Lastly, we also investigate word order in complex clauses, focusing on reported speech in which the order of the complement clause depends on the placement of a speech verb 'to say'. In Chinese languages with the verb-medial basic word order, the complement clause always follows the speech verb, while the verb-final Chinese and Tibeto-Burman languages naturally place the complement clause before the speech verb, as contrasted in (40) and (41).

**Dancun Taishanese**

- (40) *nɪ<sup>33</sup> kɔŋ<sup>55</sup> k'ui<sup>33</sup> hieŋ<sup>33</sup> ŋɔi<sup>33</sup> tiu<sup>32</sup> lɔi<sup>22</sup> lɔ<sup>55</sup>*  
 2SG say 3SG listen [1SG at.once come PART]  
 'You tell him that I am coming right away.' (Yue-Hashimoto 2005, p. 392)

**Tangam (Macro-Tani)**

- (41) *ami=de kutuk=de to-ma(ŋ)-ne=do en-la keta-duŋ*  
 person=ANAP [frog=ANAP exist-NEG-NMLZ:SUBJ=QUOT] say-NF look-IPFV  
 'The man looked [into the hole in the tree], thinking that the frog might be in there.'  
 (Post 2017, p. 125)

At the same time, there are also verb-final languages such as Bangru, which besides the preverbal complement clause (42a) also allows the postverbal complement clause (42b).

**Bangru (Miji)**

- (42) a. *madhu ravi-ya miavi teacher té-ro*  
 Madhu [Raviya-ACC good teacher] say-PST  
 'Madhu said that Ravi is a good teacher.'  
 b. *mari té-ro iya-ga mania čo mia engineer-ro*  
 Mary say-PST [3SG-GEN mother DET good engineer-PST]  
 'Mary said that her mother is a good engineer.'  
 c. *madhu ravi-ya bajar liadi ka té-ro*  
 Madhu Raviya-ACC [market go to] say-PST  
 'Madhu told Raviya to go to the market.' (Devi and Ramya 2017, p. 48)

Interestingly, when the complement part is built on a non-finite verb, it always precedes the speech verb ‘to say’ as in (42c), revealing the original head-final pattern at a deeper syntactic level (see [Devi and Ramya 2017](#), p. 48, and further discussion on subordinate clauses in Section 5.2).

## 5. Discussion of Methodological Issues in the Reconstruction of Protolanguage Syntax

This section aims to highlight four main issues arising from the present study which can contribute to a more general understanding and discussion of reconstructing and explaining syntactic changes. Visualisation of quantified data will also be presented to support arguments made on the basis of the data discussed hitherto.

### 5.1. Internal Explanation for the Head-Initial Noun Phrase Structures: Adjectives, Gender Specifiers and Quantifiers

The investigation of word order in noun modification has shown deviation among different constructions in Chinese and Tibeto-Burman languages. On the one hand, modifying nouns and possessors are consistently placed before head nouns, maintaining a canonical head-final pattern in Sino-Tibetan at deeper syntactic levels. On the other hand, some other modifiers such as quantifiers and adjectives are consistently prenominal in Chinese but more frequently postnominal in other Tibeto-Burman languages contrary to the Greenbergian word order correlations expected from verb-final languages. The reason for the latter may be due to a similar mechanism which is responsible for head-initial structures emerging in southern Chinese languages as discussed in Section 4.1. Therefore, we are particularly interested in the features which show deviation from expected typology, i.e., adjectives, gender specifiers and quantifiers. In this section, we will use evidence from Chinese languages as an explanatory tool for what could have happened in Tibeto-Burman at much earlier stages.

In Sino-Tibetan languages, adjectives syntactically behave like verbs, from which they have diachronically developed ([Genetti 2011](#), pp. 181–82). In terms of paradigmatic syncretism, we can observe for instance the co-expression of the unstressed linking words in Mandarin *de* 的 and Hui *ka* and the Cantonese classifier *ge*<sup>3</sup> 嘅 for different types of noun modifications: possessor, adjective and relative clause. Regarding adjectives, there is evidence that several languages morphologically mark adjectives with dedicated derivational affixes when adjectives are used in a non-canonical postnominal position. This phenomenon is particularly observed in colour modifiers such as in Bai *xo*<sup>55</sup> *ts*<sup>h</sup> *ε*<sup>44</sup> -*no*<sup>33</sup> [flower red-ATTR] ‘a red flower’ and *se*<sup>44</sup> *lv*<sup>44</sup> -*no*<sup>33</sup> [leaf green-ATTR] ‘green leaf’ ([Wang 2005](#), p. 108) and in Tujia *ka*<sup>1</sup> *pu*<sup>1</sup> *mian*<sup>1</sup> -*jie*<sup>1</sup> [flower red-ATTR] ‘red flower’ ([Brassett et al. 2006](#), p. 101). It might be possible that colour terms, which are frequent in use, was one potential source construction from which a language may start to treat the adjectival constituent as a head modified by the preceding noun, i.e., ‘the flower red’ and ‘the leaf green’.

The same principle can be applied to gender specifiers. [Wu \(2005, p. 106\)](#) is of the opinion that both orders stem from the Archaic Chinese head-final pattern, but their distribution varies across animal referents, and the head-initial pattern is a result of interpreting the gender-specifying constituent as a head noun. Consider, for instance, the latter constituent of Changsha Xiang example *po*<sup>13</sup> *tsi* ‘old woman; grandmother’ in (14c). A similar explanation is also given for Caijia postnominal gender specifiers, e.g., *tɕie*<sup>55</sup> ‘sister’ ([Lü 2022](#), p. 56). This interpretation would ultimately be comparable to how we treat other postnominal markers, such as diminutive and plural markers. The use of the term ‘child’ as a diminutive marker has been reported for Sino-Tibetan and also cross-linguistically, e.g., Hakka *tyau*<sup>1</sup> *tsai*<sup>3</sup> [bird DIM] ‘little bird’ in which the diminutive constituent 仔 *tsai*<sup>3</sup> etymologically derives from 子 ‘child’. Beyond Sino-Tibetan, there has been a discussion of other East Asian agglutinative languages, in which plural markers have been grammaticalised from a collective noun, e.g., the Japanese postnominal plural marker *tachi* from Old Japanese *tati* has been reconstructed to a Pre-Old Japanese noun *\*totwi* (< Pre-Proto-Japanese *\*tatəj*) with a lexical meaning ‘fellow, everyone, together’ ([Francis-Ratte 2021](#)).

As for quantifiers, the head-initial pattern in Tibeto-Burman could be an analogy to postverbal quantifiers used in the comparative construction in Chinese, consider the adverbial *san*<sup>33</sup> *li*<sup>13</sup> *mi*<sup>41</sup> [three centimetre] ‘three centimetres’ in (24b). However, apart from Karenic, Mruic and Caijia, we found no evidence from other contemporary verb-final Tibeto-Burman languages which would allow the insertion of postverbal adverbials. We maintain this as one possible explanation which requires more evidence from future studies.

The current view would then assume that northern Chinese languages with consistent prenominal modification must have undergone some reinforcement of the rigid head-final pattern for noun phrase structures at some point in their history (see also a similar case for Zhoutun comparative constructions discussed in Zhou 2023). The hypothesis is in line with the Altaicisation of Chinese which mainly concerns the northern part of Chinese speaking areas (as discussed in Hashimoto 1976; Xu 2006; Janhunen 2007, 2012; Szeto et al. 2018; Szeto 2019; Szeto and Yurayong 2022). The reason is that languages of the Altaic type are consistently head-final with close to zero tendency of realizing head-initial structures, in terms of the features given in Table 3.

Given the previous discussion, we have good reason to reconstruct Proto-Chinese as well as Proto-Sino-Tibetan noun phrases as being consistently head-final. Firstly, the head-initial deviants in Tibeto-Burman can be explained to a certain degree by an internal development through reanalysis of modifier constituents. Secondly, the predominance of head-final noun phrase structures in Chinese, particularly the northern varieties, may also have been reinforced by contact with the consistently head-final Altaic languages.

## 5.2. Archaism in the Head-Final Subordinate Clauses: Reported Speech and Negation

Compared to noun phrases, verb phrases are more hybrid and vary across Chinese languages. Crucially, the head-initial deviants have been attested since the Archaic Chinese stage (as discussed in Section 2.3), so the directionality shift to head-initial verb structures must have been well established and can be considered a relatively early Proto-Chinese innovation. In this section, we proceed beyond Proto-Chinese to other Tibeto-Burman languages. As noted in Section 4.3, some languages may organise reported speech and negative construction word order asymmetrically between main and subordinate clauses. Such asymmetry has been discussed in the typological literature (see Bybee 2001 as discussed in Section 2.1), and it seems to be applicable to the Sino-Tibetan context to a certain extent.

In (42), we have observed word order variation in Bangru reported speech with a finite compliment clause, which allows both a preverbal (42a) and postverbal position (42b). However, a preverbal complement clause as in (42c) is the only acceptable pattern when the verb is non-finite. Similarly, negative constructions may include both preverbal and postverbal negators in main clauses, as in the Lepcha example (39). Yet, few languages of the circumflexive negator type drop the postverbal negator in subordinate clauses because the postverbal slot is occupied, for instance, by a linker in conditional clauses or a relativiser in relative clauses, as is the case for Sgaw Karen (Kerbs Forthcoming) and Burmese in (43).

### Burmese (Burmo-Qiangic)

- (43) a. *paiʔsʰan*                      *mə-ɕí-bù*  
          money                      [NEG-exist-NEG]  
          ‘(I) don’t have money.’
- b. *paiʔsʰan*                      *mə-ɕí-yin*                      *mə-la-né*                      *le*  
          money                      [NEG-exist]-if                      NEG-COME-PROH                      EMPH  
          ‘If you don’t have any money, don’t come!’ (Jenny and Hnin Tun 2016, p. 360)

The opposite direction is reported in Eastern Pwo Karen where main clauses employ the postverbal negator ʔé (44a) whereas subordinate clauses have circumflexive negators (44b).

**Eastern Pwo Karen (Karenic)**

- (44) a. *ʔəwê ʔán bá mì dài ʔé*  
 3sg eat opp rice still NEG  
 ‘He has not managed to eat rice yet.’ (Kato 2021, p. 356)
- b. *ʔəwê lə yé lá jò bá ʔəkhúcònjə bá mà*  
 3sg NEG come loc here NEG because 1sg must do  
 ‘Because he did not come here, I have to do.’ (Kato 2021, p. 357)

In some other languages, the construction may remain identical to main clauses but there can be some morphophonological alternation of the postverbal negator, as in Phedappe Limbu *mɛN-V-nɛn-n* vs. *mɛN-V-mna*, and East-Central Tangkhul Naga *ə-V-mə* vs. *ə-V-k<sup>h</sup>ə-mo*, respectively.

In the typology of negation, the former type in Burmese and Karenic is labelled as “constructional asymmetry”, and the latter type in Phedappe Limbu and East-Central Tankhul is labelled as “paradigmatic asymmetry” (see Miestamo 2005, 2013). The asymmetry between main and subordinate clauses in negative constructions is, however, not a widely observed phenomenon in Sino-Tibetan languages according to our observation. Moreover, the postverbal element of circumflexive negators can be etymologised as deriving from a negative status marker because the preverbal *\*ma-* is reconstructed as a Proto-Tibeto-Burman standard negator alongside the imperative *\*ta-/da-* (Benedict 1972, p. 97; Matisoff 2003, p. 162). Thus, the triggering effect of clausal complexity does not seem to go back to Proto-Sino-Tibetan, nor to any second level protolanguages. In any case, the similarity between Burmese and Sgaw Karen could stem from more recent contact with Burmese-speaking ruling elites in many Karenic speaking areas in the past centuries (see also Kato 2019).

Another interesting observation comes from Raji which uses a postverbal *-ma* as a standard negator as in (45a), but its position shifts to preverbal in non-finite verb structures, such as adverbial (45b), complement (45c) and relative clauses (45d).

**Raji (Raji-Raute)**

- (45) a. *ŋa bətaŋ dza-tɕ-ma*  
 1sg rice eat-PFV-NEG  
 ‘I have not eaten rice.’ (Khatri 2012, p. 82)
- b. *daktar samai-fia ma-tonpi-tinŋ beramya si-k-a*  
 doctor time-LOC NEG-arrive-NF patient die-same.day-PST  
 ‘The patient died since the doctor did not come on time.’ (Khatri 2012, p. 83)
- c. *ma-dzak-tinŋ noknŋ yo-k-i*  
 NEG-put-NF good be-same.day-NPST  
 ‘Is it good not to put?’ (Khatri 2012, p. 83)
- d. *ma-das-tinŋ aŋe aŋe*  
 NEG-know-NF what what  
 ‘Unskilled people do not know.’ (Khatri 2012, p. 83)

Applying Bybee’s (2001) idea that subordinate clauses preserve original constructions better than in main clauses, we can assume that the original position of the Raji negator was preverbal, in line with the closely related Rawat preverbal negator *hã-*.

As a diachronic implication, subordinate clauses seem to reveal the original syntactic patterns of reported speech and negation as they conform to the general head-final tendency across the family. The previous analysis, thus, supports the view that subordinate clauses tend to preserve original syntax better than in main clauses and are suitable for being used as a criterion for reconstructing protosyntax (as discussed in Bybee 2001; Jenny 2015, 2020, and in Section 2.1). Crucially, this criterion can also be effectively used to describe the variation and deviating patterns of word order observed in other lesser studied Chinese languages with heavy contact influence from neighbouring languages, such as Dungan in Central Asia and Junhua and Maihua in Southeast Asia.



### 5.3. Reconstructing Non-Harmonious Word Order Patterns for Proto-Chinese

The data discussed hitherto have shown that none of the Chinese and Tibeto-Burman languages are fully head-final, i.e., receiving the full score in Supplementary Material B. In any case, such a hybrid profile of Chinese word order is not to be used as an argument for the mixed origin of Proto-Chinese (Benedict 1972; Bodman 1980; DeLancey 2013), and as discussed in Sections 2.3, 4.3 and 5.1, this has crucially also provided an explanation from the perspective of language-internal change (à Thomason 2010). At least some of the head-initial deviants can be explained by internal reconstruction through grammaticalisation, particularly in the case of adpositional phrases, whereas certain other constructions still need further analysis in the framework of historical syntax (a challenge previously encountered also in Dryer 2003).

At a methodological level, we propose an approach to deal with the varying degrees of word order variation across Sino-Tibetan and their development by referring to the notion of “protolanguage”. Namely, any protolanguage which has been discussed in historical linguistics dates no further than ca. 6000–8000 years back in time, which is significantly much later than the emergence of human speech ability a hundred thousand years ago (e.g., Oller 2000; Lieberman 2007; Perreault and Mathew 2012). This thereby leaves some room for the possibility of reconstructing a protolanguage with hybrid word order profiles which conform with one of the scenarios given in Table 1. This seems to be the case for the Proto-Chinese intermediate stage and the earlier Proto-Sino-Tibetan stage. The approach would be in line with a functional approach to word order variation in a tendency-like manner rather than an invariant parameter (as discussed in Section 2.1). With this starting point, we will discuss further how the hybrid profile of Proto-Chinese and Proto-Sino-Tibetan should be analysed.

Regardless of variation at the level of individual languages, the head-final baseline is in place for the majority of the sample Chinese and Tibeto-Burman languages. A deviation from the general tendency described above occurs in Karenic languages which receive notably lower than half of the head-finalness score in Supplementary Material B. This supports the view of an early radical typological change in Proto-Karenic, considered to be the result of intense and long-standing contact with Mainland Southeast Asian verb-medial languages, most notably with Austroasiatic languages, Monic and then Palaungic languages and more recently with Tai languages (see Matisoff 2000; Manson 2009; and Kato 2019). The same can be said for Chinese (excluding modern contact varieties in the northwest) as there is a geographically significant cline along which the degree of head-finalness decreases towards the south, an area where most contact languages prefer the head-initial structures, similar to the Northern Eurasian cline observed in Nichols (1992) (see Figure 6 in Section 5.4).

This observation also gives an implication that change of head directionality does not necessarily concern the entire word order system. Instead, the shift of constituent orders gradually becomes activated in individual constructions in a predictable way (such as the hierarchy of noun phrase modification and the postulation of shift order discussed in Section 2.1). For instance, the relational order between adjectives and demonstratives observed from Sino-Tibetan languages contradicts implicational universals and the postulation of word order shift (as discussed in Section 2.1), namely that adjectives tend to change before demonstratives. As we see from the counterexample (17), Toto retains the prenominal position of demonstratives more firmly than that of adjectives.

In any case, the implicational universals, according to which relative clauses would change their directionality last are supported by a report about demonstratives in Daai Chin, in which prenominal demonstratives cannot occur adjacent to a relative clause or possessor as a modifier in the noun phrase, since demonstratives would otherwise be pushed to the postnominal position (So-Hartmann 2009, pp. 133, 141). A similar complementary relation between prenominal relative clauses and postnominal demonstratives has been observed in Rawat (Rastogi 2012, p. 86). Among Chinese languages, Tunxi Hui noun relativization, with the demonstrative following the relative clause and preceding the head

noun, as in (18b), is considered indigenous (Lu 2018, p. 168). The same pattern is also reported for Yichun Gan (Li 2018, p. 250).

With the quantitative data from Sino-Tibetan languages presented in Supplementary Material B, in Table 4 and the discussion to this point, we propose in Table 5 an alternative schema to the model in Table 1, in which noun modification types have shifted from head-final to head-initial in a chronological order.

**Table 5.** Hierarchy of directionality change in word order of nominal structures.

Relative Chronology	The Version Based on Implicational Universals in Hawkins (1983, p. 75), Dryer (1992b) and Croft (2003, p. 123)	The Version Based on the Sino-Tibetan Data
First		Gender specifiers
	Quantifiers	Quantifiers
	Demonstratives	Adjectives
	Adjectives	Demonstratives
	Possessors	Relative clauses
	Relative clauses	Compounds
Last		Possessors

Based on our data and proposal, possessors and then compounds would be at even deeper syntactic levels of the noun phrase and thereby more resistant to change than relative clauses. At the same time, gender specifiers and then quantifiers are clearly more mobile as is seen in the ongoing reanalysis of gender-specifying morphemes as head nouns (as discussed in Section 5.1) and the word order variation of quantifiers used as an abstract measurement in comparatives, as shown in (24b) and discussed in Section 4.1.

Applying this hierarchy to the comparative Sino-Tibetan data discussed in the present study, we reconstruct head-final word order profile for the Proto-Sino-Tibetan noun phrase structures which are retained in Proto-Chinese. Our assumption as shown in Table 6 is that other intermediate protolanguages later gradually underwent the shift towards head-initial structures in quantifiers, gender specifiers and adjectives in a similar manner to what has recently taken place in southern Chinese languages, as discussed in Section 5.1. The opposite direction with Chinese being innovative in restructuring the noun phrase structures into a rigid head-final type would be less probable. Firstly, the secondary reinforcement of head-final tendency under Altaic contact (as discussed in Section 5.1) would not have taken place and had significant effects on Chinese typology until the second half of the first millennium CE. Secondly, there has been hitherto no information about any prehistoric head-final languages to the west and north of Chinese speaking areas other than the cognate Tibeto-Burman and the Altaic type of languages in the north, which could have been candidates for the pre-Altaic restructuring stage.

**Table 6.** A reconstruction of Proto-Chinese and Proto-Sino-Tibetan noun phrases: N = Noun; N<sub>[H]</sub> = Head noun; Quant = Quantifier; Gend = Gender specifier; Adj = Adjective; Dem = Demonstrative; Rel = Relative clause; and Poss = Possessive.

	Head-Initial Structures	Head-Final Structures
Other intermediate protolanguages	NQuant & NGen & NAdj	DemN & RelN & NN <sub>[H]</sub> & PossN
Proto-Chinese		
Proto-Sino-Tibetan	QuantN & GendN & AdjN & DemN & RelN & NN <sub>[H]</sub> & PossN	

The proposed hierarchy can be further used to define the complexity level of different noun modifications of Sino-Tibetan languages typologically (as discussed in Section 5.2). A similar analysis can be empirically tested with other language families to achieve ultimately a cross-linguistically valid prediction of change in word order patterns.

In contrast, Proto-Sino-Tibetan verb phrase structures must have been consistently head-final, which is largely retained in Tibeto-Burman, apart from Karenic affected by contact with Mainland Southeast Asian languages and Burmo-Qiangic, and Macro-Bai with a recent restructuring stage under Mandarin influence. At the same time, Proto-Chinese verb phrases presumably shifted their pattern to head-initial in the constructions indicated in Table 7, which might have been the result of pre-historic contact with Mainland South-east Asian languages (as hypothesised in DeLancey 2013).

**Table 7.** A reconstruction of Proto-Chinese and Proto-Sino-Tibetan verb phrases: V = Verb; O = Direct object; Quot = Quotation; Adj = Adjective; Comp = Standard of comparison; Loc = Locational adverb.

	Head-Initial Structures	Head-Final Structures
Proto-Chinese	VO & VQuot & AdjComp	LocV
Pre-Proto-Chinese & Other intermediate protolanguages	OV & QuotV & CompAdj & LocV	
Proto-Sino-Tibetan		

Among these verb phrases, the comparative, existential and reported speech constructions seem to have deviated from other Sino-Tibetan languages and shifted to the head-initial pattern early enough to become stabilised in Archaic Chinese, similarly to the grammaticalisation of location and directional verbs into prepositions. At the same time, the location adverbial remains largely preverbal. The situation after intense contact with head-final Altaic languages then presumably led to the reinforcement and secondary re-establishment of pre-Proto-Chinese/Proto-Sino-Tibetan head-final construction types. Meanwhile, the situation with ditransitives, which is a pragmatically highly variable construction as discussed in Section 2.3 in connection to (7) and (9), remains to be studied further.

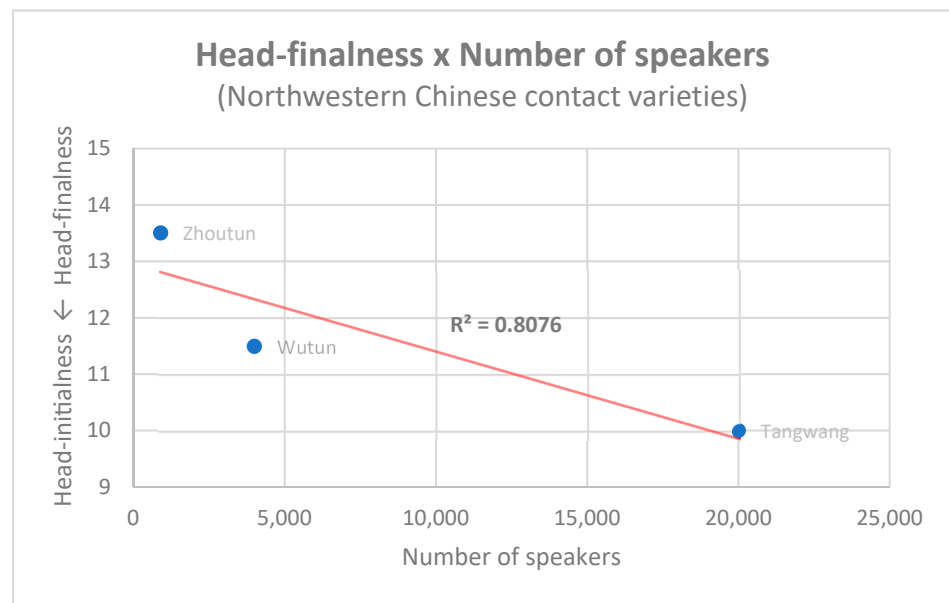
#### 5.4. Statistical Correlations between Word Order Variation and Sociological Factors

To complement the internal and external reconstruction of Proto-Chinese and Proto-Sino-Tibetan syntax proposed throughout Section 5, we also conduct coefficient of determination ( $R^2$ , Wright 1921) tests to explore whether extralinguistic factors can also add significant explanatory power. The degree of significance is determined by the following scalar: 0.00–0.19 (not interesting) → 0.20–0.49 (low) → 0.50–0.79 (moderate) → 0.80–1.00 (high). As a precaution, the correlations discussed in this section are predominantly statistical and do not necessarily indicate causality between the tested variables.

In terms of the speech community, language change can be affected by the proportion of L1 and L2 speakers and by linguistic awareness, as well as by geographical factors (as discussed in Section 2.4). However, the information concerning the number of L2 speakers is not always reported, let alone the accuracy of the number of overall speakers of a language. Likewise, it is often difficult to assess the linguistic awareness of a given speech community. We acknowledge the significance of all the aforementioned factors for the explanation of language change, but due to the limited amount of available data, we limit our scope in this section to two extralinguistic factors: (1) the overall number of speakers (Scenario type 1) and (2) geographical location of a speech community (Scenario types 2 and 3). Then, we run ca. 30 different test scenarios to detect statistical correlations between the extralinguistic factors and two aspects of word order: (1) the degree of head-final tendency and (2) the degree of hybrid word order profile. Due to the limitations of space, we present in this section only scenarios which show statistically interesting results.

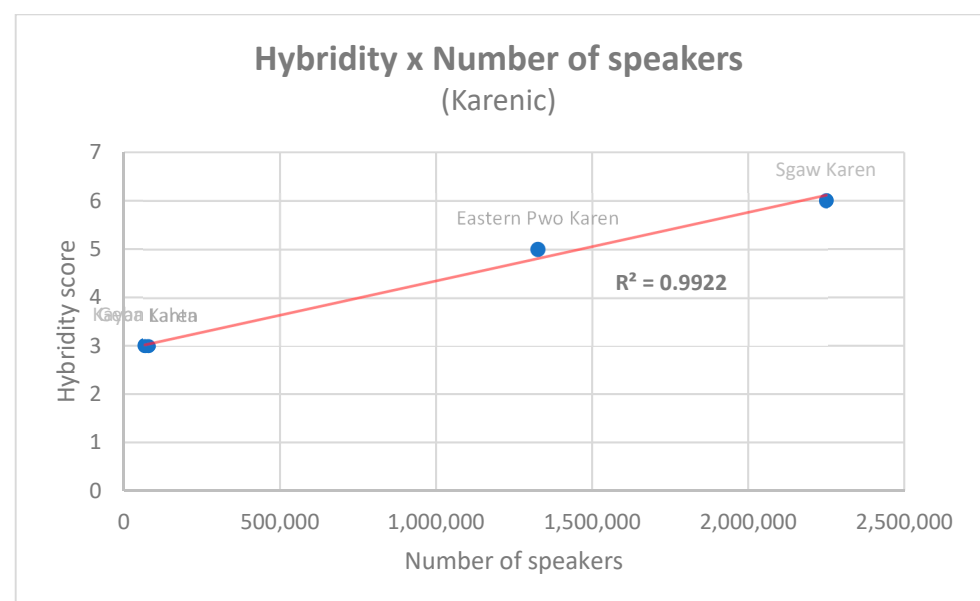
*Scenario type 1: Head-finalness x Number of speakers*

At a macro level, there is no interesting result ( $R^2 < 0.01$ ) across the entire Sino-Tibetan family regarding the statistical correlation between head-finalness and number of speakers. However, when we select a filtered sampling with only languages which share a similar scenario of historical development, we detect significant correlations among northwestern Chinese contact varieties (Tangwang, Wutun, Zhoutun) and Karenic languages. For the former case, the head-final tendency decreases as the number of speakers increases in a relatively significant fashion ( $R^2 \approx 0.81$ ), as is shown in Figure 3.



**Figure 3.** The statistical correlation between head-final tendency and number of speakers across northwestern Chinese contact varieties.

At the same time, the number of Karenic speakers statistically correlates in a highly significant manner ( $R^2 \approx 0.99$ ) with the number of hybrid constructions, i.e., variation in head directionality (assigned as 0.5. in Supplementary Material B), as is observable in Figure 4.

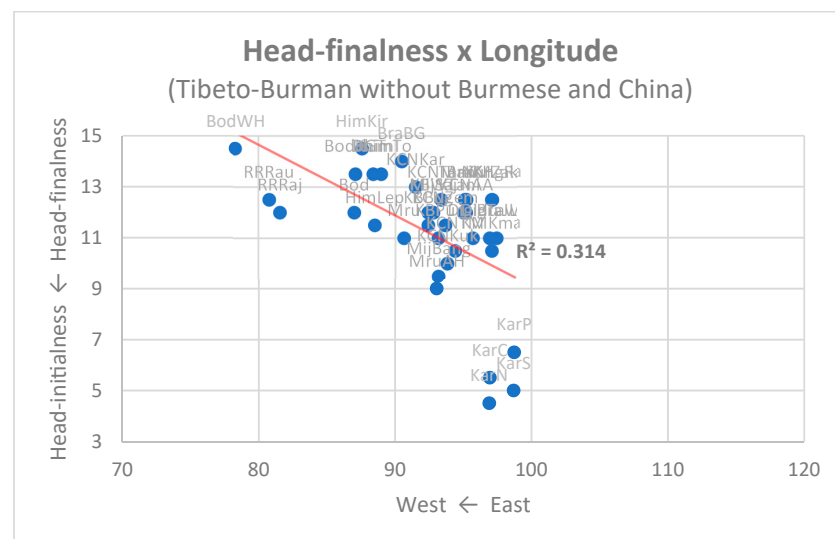


**Figure 4.** The statistical correlation between the number of hybrid constructions and number of speakers across Karenic.

Both micro-area cases statistically imply that the more speakers a speech community has, the more a language tends to tolerate variation and hybridity, from which it likewise may become more deviant from its baseline syntactic pattern (head-final in this case). However, we speculate that the statistical correlation between the number of speakers and syntactic variation does not work independently without consideration of other sociolinguistic factors. The reason we detect the micro-areal correlations in Figures 3 and 4 could be due to the fact that these languages belong to the same genealogical groups, which makes the tested scenarios valid in their own sociohistorical contexts.

#### Scenario type 2: Head-finalness x Longitude

Replicating the macro-area investigation by Nichols (1992) in the Sino-Tibetan context, we notice that Chinese languages are too deviant from the rest and affect the result, which yields no interesting statistical significance ( $R^2 \approx 0.11$ ). Among Chinese languages themselves, there is also only an uninteresting significance detected ( $R^2 \approx 0.08$ ). However, when we conduct a separate test for Tibeto-Burman, the correlation is improved ( $R^2 \approx 0.23$ ). We also try scenarios which exclude languages considered as outliers: (1) Burmese which arrived in its current area less than a millennium ago and (2) languages spoken in China (Na-Qiangic, Tujia, Baic and Tai-Li) as they are reportedly undergoing a heavy influence from standard or regional Mandarin syntax. With this filter, we achieve a better result ( $R^2 \approx 0.31$ ), as shown in Figure 5.



**Figure 5.** The statistical correlation between the head-final tendency and longitude across Tibeto-Burman, excluding Burmese and the languages spoken in China.

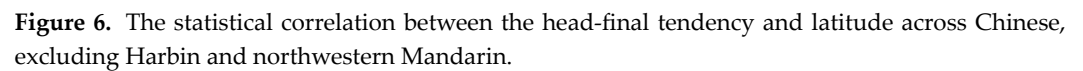
This scenario type does not indicate a high significance in terms of intralinguistic and extralinguistic correlations. In any case, we maintain that the east–west cline is still there in the Sino-Tibetan sphere (as observed for individual features in Dryer 2003) and might have decreased due to its time depth which could go back to the early migration and expansion of the Sino-Tibetan speaking populations, similarly to the east–west cline observed across Northern Eurasia and North America by Nichols (1992). Such a prehistoric cline where the west retains the head-final tendency more firmly seems to be due to the prehistoric Trans-Himalayan migration from the Sino-Tibetan *Urheimat* located further to the east (see also discussion on the Sino-Tibetan *Urheimat* in Peiros 1998, p. 217; van Driem 2007; Sagart et al. 2019; Zhang et al. 2019).

#### Scenario type 3: Head-finalness x Latitude

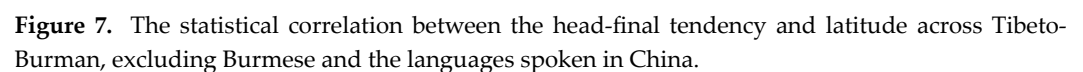
In another axis, we identify the cline from north to south by conducting two separate tests for Chinese and Tibeto-Burman, as the correlation rate across the entire family is relatively low ( $R^2 \approx 0.17$ ) and might have been affected by many layers of later con-



For Chinese, the north-to-south cline has a relatively high degree of significance ( $R^2 \approx 0.8$ ), as shown in Figure 6.



Tibeto-Burman, meanwhile, shows a lower but still moderately significant degree of correlation ( $R^2 \approx 0.61$ ), as shown in Figure 7.



Similar to the case of Chinese above, the cline in Figure 7 might also give an areal implication of the typological distinction between core Tibeto-Burman and Mainland Southeast Asian type of Tibeto-Burman, including Karenic, Kuki-Chin and Mruic.

The observation based on latitude shows that the north-to-south cline (as hypothesised in Section 2.4) is applicable to the Sino-Tibetan family, and further studies can retest this hypothesis with a larger sampling which includes also other language families of Eastern Eurasia.

## 6. Conclusions

In the present study, we have investigated word order of nominal and verbal structures in Sino-Tibetan languages in terms of head-final vs. head-initial patterns by using evidence from both internal reconstruction and external factors, such as language contact, geographical location and sociolinguistic context. We have shown that while the Sino-Tibetan language family in general tends to favour head-final structures; individual languages show considerable degrees of variation in the number of head-initial structures allowed for. We have discovered that internal development through reanalysis and contact with neighbouring languages are both relevant for explaining variation and changes in word order structures. It is also evident from the results of this study that nominal and verbal categories might be non-harmonious in one and the same language, as has been the case for Chinese since the Proto-Chinese stage.

In terms of extralinguistic variables, we have found that a sociolinguistic factor (number of speakers) and geographical factors (latitude and longitude) show statistical correlations with word order variation. At least on the level of micro-areas, the discussed extralinguistic factors can ultimately also give some implications on directionality changes in word order of Sino-Tibetan language communities. Crucially, we have demonstrated that in the case of modern Chinese contact varieties of Northwest China, smaller speech communities tend to be more consistent with the head-final tendency. Among Karenic languages, meanwhile, larger speech communities tend to tolerate more variation in word order patterns, which, in turn, causes deviation from the head-final syntactic baseline. The most significant factor on the level of the entire language family seems to be latitude, which statistically correlates with the degree of head-final tendency, and therefore the north–south cline is of considerable significance for Sino-Tibetan languages. At the same time, the east–west cline is in line with the hypothesis about the early Trans-Himalayan migration from the east. In any case, the role of the proportion of L2 speakers, as well as language attitudes and language awareness of the speakers as explanations to language change in the Sino-Tibetan family, still requires further data and research.

In the current study, we have used the variety sampling method with stricter areal and genealogical controls for language selection which yields results as discussed throughout the entire work. In future studies, the application of a larger sample of languages and typological features as well as more versatile extralinguistic factors related to language sociology can shed light on variation across Sino-Tibetan at an even more fine-grained level.

**Supplementary Materials:** The Supplementary Materials A and B can be downloaded at: <https://www.mdpi.com/article/10.3390/languages8020112/s1>.

**Author Contributions:** Conceptualization, C.Y. and E.S.; Methodology, C.Y.; Formal analysis, C.Y. and E.S.; Investigation, C.Y. and E.S.; Resources, C.Y. and E.S.; Writing—original draft, C.Y.; Writing—review & editing, E.S.; Visualization, C.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was part of the project “Negation in Clause Combining: Typological and Usage-Based Perspectives (NiCC)” funded by the Academy of Finland grant number 332529; and the project “Evidentiality in social interaction: a cross-linguistic perspective” funded by Kone Foundation grant number 0213537-1.

**Institutional Review Board Statement:** The study was exempted from ethical review by the Committee for Research Ethics (Social Sciences), Faculty of Social Sciences and Humanities, Mahidol University (2022/001.1701, 17 January 2022).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available in Supplementary Materials A and B.

**Acknowledgments:** This research was initiated through the project “Northeast and Southeast Asian Studies Network in Finland and Thailand (NSEANET)” which was funded by the Finnish National Agency for Education (EDUFI) during 2020–2023. Under this cooperation, the two authors co-taught the course “Chinese Linguistics” in autumn 2021 for the Research Institute for Languages and Cultures, Mahidol University, and the Department of Languages, Faculty of Arts, University of Helsinki, which served as a catalyst for this research. We wish to express our gratitude to the students who participated in this course, as well as the students who were enrolled in the course “Southeast Asian Linguistic” at Mahidol University in spring 2022, for their contributions to the data collection as acknowledged in Supplementary Material A and for sharing their interesting observations on word order variation across Sino-Tibetan languages. We also would like to extend our gratitude to Kaius Sinnemäki for his fruitful consultation on the sociotypological approach, which has greatly benefited the theoretical contextualisation of this research. Moreover, we are enormously thankful to Jonathan Banks for his exceptional proofreading and editing of the English language. The usual disclaimer applies. Open access funding was provided by the University of Helsinki.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

1	First person	LOC	Locative
2	Second person	MASC	Masculine
3	Third person	NEG	Negation
ACC	Accusative	NF	Non-finite
AGENT	Agent	NOM	Nominative
ANAP	Anaphoric	NMLZ	Nominaliser
APPL	Applicative	NPST	Non-past
ASP	Aspect	OBJ	Object
ATTR	Attributive	OPP	Opportunity
CLF	Classifier	PART	Particle
COMP	Comparative	PFV	Perfective
COP	Copula	PL	Plural
DAT	Dative	POSS	Possessive
DECL	Declarative	PROH	Prohibitive
DET	Determiner	PST	Past
EMPH	Emphatic	PTCP	Participle
EXCL	Exclamation	QUOT	Quotative
FEM	Feminine	REAL	Realis
GEN	Genitive	RECIP	Reciprocal
HAB	Habitual	REFL	Reflexive
INCOMPL	Incompletive	REL	Relative clause
INTJ	Interjection	RES	Resultative
IPFV	Imperfective	SG	Singular
LINK	Linker	SUBJ	Subject

## References

- Aldridge, Edith. 2013. Survey of Chinese Historical Syntax Part I: Pre-Archaic and Archaic Chinese. *Language and Linguistics Compass* 7: 39–57. [\[CrossRef\]](#)
- Aldridge, Edith. 2015a. Old Chinese Syntax: Basic Word Order. In *Encyclopedia of Chinese Language and Linguistics*. Edited by Rint Sybesma. Leiden: Brill.
- Aldridge, Edith. 2015b. Pronominal Object Shift in Archaic Chinese. In *Syntax over Time: Lexical, Morphological and Information-Structural Interactions*. Edited by Theresa Biberauer and George Walkden. Oxford: Oxford University Press, pp. 350–70.

- Ansaldo, Umberto. 1999. Surpass Comparatives in Sinitic: Typology and Grammaticalization. Doctoral dissertation, Stockholm University, Stockholm, Sweden.
- Ansaldo, Umberto. 2010. Surpass comparatives in Sinitic and beyond: Typology and grammaticalization. *Linguistics* 48: 919–50. [CrossRef]
- Arcodia, Giorgio Francesco. 2007. Chinese: A language of compound words. In *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*. Somerville: Cascadilla Proceedings Project, pp. 79–90.
- Basumatary, Chibiram. 2016. A Descriptive Grammar of Toto. Doctoral dissertation, Assam University, Silchar, India.
- Benedict, Paul K. 1972. *Sino-Tibetan: A Conspectus*. Cambridge: Cambridge University Press.
- Bennett, Paul. 1979. A critique of the Altaicization hypothesis. *Cahiers de Linguistique Asie Orientale* 6: 91–104. [CrossRef]
- Bentz, Christian, and Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3: 1–27. [CrossRef]
- Bickel, Balthasar. 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In *The Oxford Handbook of Linguistic Analysis*, 2nd ed. Edited by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press, pp. 901–23.
- Bodman, Nicholas C. 1980. Proto-Chinese and Sino-Tibetan: Data towards establishing the nature of the relationship. *Cornell Linguistic Contribution* 3: 34–199.
- Brassett, Cecilia, Philip Brassett, and Meiyang Lu. 2006. *The Tujia Language*. München: LINCOM Europa.
- Burling, Robbins. 2003. *The language of the Modhupur Mandi, Garo: Vol. I: Grammar*. Ann Arbor: University of Michigan.
- Bybee, Joan L. 2001. Main clauses are innovative, subordinate clauses are conservative. In *Complex Sentences in Grammar and Discourse: Essays in honor of Sandra A. Thompson*. Edited by Joan L. Bybee and Michael Noonan. Amsterdam: John Benjamins, pp. 1–17.
- Cao, Zhiyun, ed. 2008. *Hànyǔ fāngyán dìtújí [Linguistic Atlas of Chinese Dialects]*. Beijing: The Commercial Press.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chappell, Hilary M., ed. 2001. Synchrony and diachrony of Sinitic languages: A brief history of Chinese dialects. In *Sinitic Grammar: Synchronic and Diachronic Perspectives*. Oxford: Oxford University Press, pp. 3–28.
- Chappell, Hilary M., ed. 2015. Linguistic areas in China for differential object marking, passive, and comparative constructions. In *Diversity in Sinitic Languages*. Oxford: Oxford University Press, pp. 13–52.
- Chappell, Hilary M., Li Ming, and Alain Peyraube. 2007. Chinese linguistics and typology: The state of the art. *Linguistic Typology* 11: 187–211. [CrossRef]
- Chen, Weirong. 2020. *A Grammar of Southern Min: The Hui'an Dialect*. Berlin: Mouton de Gruyter.
- Chomsky, Noam, and Howard Lasnik. 1993. The theory of principles and parameters. In *Syntax ein Internationales Handbuch Zeitgenössischer Forschung*. Edited by Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann. Berlin: Mouton de Gruyter, vol. 1, pp. 506–69.
- Croft, William. 2003. *Typology and Universals*, 2nd ed. Cambridge: Cambridge University Press.
- Dahl, Östen. 1979. Typology of sentence negation. *Linguistics* 17: 79–106. [CrossRef]
- de Sousa, Hilário. 2015. The far Southern Sinitic languages as part of Mainland Southeast Asia. In *Languages of Mainland Southeast Asia: The State of the Art*. Edited by N. J. Enfield and Bernard Comrie. Berlin: Mouton de Gruyter, pp. 356–440.
- DeLancey, Scott. 2013. The origins of Sinitic. In *Increased Empiricism: New Advances in Chinese Linguistics*. Edited by Zhuo Jing-Schmidt. Amsterdam: John Benjamins, pp. 73–99.
- DeLancey, Scott. 2014. Sociolinguistic typology in North East India: A tale of two branches. *Journal of South Asian Languages and Linguistics* 1: 59–82. [CrossRef]
- Devi, Chera, and Tarh Ramya. 2017. Bangru. In *The Languages of Arunachal Pradesh*. Edited by Lisa Lomdak. Hyderabad: Orient BlackSwan, pp. 46–64.
- Devi, Thokchom Lakhpriya. 2011. The Tangkhul Grammar of Shangshak Village. Doctoral dissertation, Manipur University, Canchipur, India.
- Devi Prasada Sastry, G. 1984. *Mishmi Grammar*. Mysore: Central Institute of Indian Languages.
- Djamouri, Redoune. 1991. Pariticles delottolon dans les inscriptions sur bronze de la dynastie des Zhou. *Cahiers de Linguistique Asie Orientale* 20: 5–76.
- Dryer, Matthew S. 1992a. The Greenbergian Word Order Correlations. *Language* 68: 81–138. [CrossRef]
- Dryer, Matthew S. 1992b. Modifying Hawkins' Prepositional Noun Modifier Hierarchy. Paper presented at the Word Order Theme Group, EUROTYP Plenary Meeting, San Sebastian, Spain, March 22–24.
- Dryer, Matthew S. 2003. Word order in Sino-Tibetan languages from a typological and geographical perspective. In *The Sino-Tibetan Languages*. Edited by Graham Thurgood and Randy J. LaPolla. London: Routledge, pp. 43–55.
- Dryer, Matthews S. 2013. Order of negative morpheme and verb. In *The World Atlas of Language Structures Online*. Edited by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *World Atlas of Language Structures Online (v2020.3) [Data set]*. Zenodo. Available online: <https://wals.info> (accessed on 6 February 2023). [CrossRef]
- Fang, Jingyun, Yongchang Song, Hongyan Liu, and Shilong Piao. 2002. Vegetation-climate relationship and its application in the division of vegetation zone in China. *Acta Botanica Sinica* 44: 1105–22.
- Feng, Shengli. 1996. Prosodically Constrained Syntactic Changes in Early Archaic Chinese. *Journal of East Asian Linguistics* 5: 323–71. [CrossRef]

- Francis-Ratte, Alexander. 2021. On the etymology of the Japanese plural suffix and its possible connection to Korean. *Asian Languages and Linguistics* 2: 24–35. [CrossRef]
- Genetti, Carol. 2011. Nominalization in Tibeto-Burman languages of the Himalayan area: A typological perspective. In *Nominalization in Asian Languages: Diachronic and Typological Perspectives*. Edited by Foong Ha Yap, Karen Grunow-Härsta and Janick Wrona. Amsterdam: John Benjamins, pp. 163–93.
- Greenberg, Joseph H. 1957. The nature and uses of linguistic typologies. *International Journal of American Linguistics* 23: 68–77. [CrossRef]
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of Language* 2: 73–113.
- Guzmán Naranjo, Matías, and Laura Becker. 2022. Statistical bias control in typology. *Linguistic Typology* 26: 605–70. [CrossRef]
- Hammarström, Harald, Robert Forkel, Martin Haspelmath, and Sebastian Bank. 2022. *Glottolog 4.6*. [Released on 24 May 2022]. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online: <http://glottolog.org> (accessed on 6 February 2023).
- Hashimoto, Mantaro. 1976. Language diffusion on the Asian continent: Problems of typological diversity in Sino-Tibetan. *Computational Analysis of Asian and African Languages* 3: 49–63.
- Hawkins, John A. 1983. *Word Order Universals*. New York: Academic Press.
- Ho, Dah-an. 1987. *Shēngyǔnxué Zhōng de Guānniàn hé Fāngfǎ [The Concept and Methods of Phonology]*. Taipei: Dahan Chubanshe.
- Huang, Yang, and Fuxiang Wu. 2018. Central Southern Guangxi as a grammaticalization area. In *New Trends in Grammaticalization and Language Change*. Edited by Sylvie Hancil, Tine Breban and José Vicente Lozano. Amsterdam: John Benjamins, pp. 105–34.
- Jackendoff, Ray S. 1977. *X-Bar Syntax: A Study in Phrase Structure*. Cambridge: MIT Press.
- Jakobson, Roman. 1958. Typological studies and their contribution to historical comparative linguistics. In *Proceedings of the Eighth International Congress of Linguists*. Edited by Eva Sivertsen. Oslo: Oslo University Press, pp. 17–35.
- Janhunen, Juha. 2007. Typological interaction in the Qinghai linguistic complex. *Studia Orientalia* 101: 85–102.
- Janhunen, Juha. 2012. On the hierarchy of structural convergence in the Amdo Sprachbund. In *Argument Structure and Grammatical Relations: A Cross-Linguistic Typology*. Edited by Pirkko Suihkonen, Bernard Comrie and Valery Solovyev. Amsterdam: John Benjamins, pp. 177–89.
- Jenny, Mathias, and San San Hnin Tun. 2016. *Burmese: A Comprehensive Grammar*. London: Routledge.
- Jenny, Mathias. 2020. Verb-Initial Structures in Austroasiatic Languages. In *Austroasiatic Syntax in Areal and Diachronic Perspective*. Edited by Mathias Jenny, Paul Sidwell and Mark Alves. Leiden: Brill, pp. 21–45.
- Jenny, Mathias. 2015. Syntactic diversity and change in Austroasiatic languages. In *Perspectives on Historical Syntax*. Edited by Carlotta Viti. Amsterdam: John Benjamins, pp. 317–40.
- Kato, Atsuhiko. 2021. Typological profile of Karenic languages. In *The Languages and Linguistics of Mainland Southeast Asia: A Comprehensive Guide*. Edited by Paul Sidwell and Mathias Jenny. London: Routledge, pp. 337–67.
- Kato, Atsuhiko. 2019. Karen and surrounding languages. *Topics in Middle Mekong Linguistics* 60: 123–50.
- Kerbs, Richard. Forthcoming. A Descriptive Grammar of Sgaw Karen. Doctoral dissertation, University of Helsinki, Helsinki, Finland.
- Khatri, Ramesh. 2008. The Structure of Verbs and Sentences of Raji. Master's thesis, Tribhuvan University, Kirtipur, Nepal.
- Khatri, Ramesh. 2012. Negativization in Raji. *Nepalese Linguistics* 27: 82–84.
- King, John Timothy. 2008. A Grammar of Dhimal. Doctoral dissertation, Universiteit Leiden, Leiden, The Netherlands.
- Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199–244. [CrossRef]
- Kurpaska, Maria. 2010. *Chinese Language(s): A Look through the Prism of the Great Dictionary of Modern Chinese Dialects*. Berlin: Mouton de Gruyter.
- Ladd, D. Robert, Seán G. Roberts, and Dan Dediu. 2015. Correlational studies in typological and historical linguistics. *Annual Review of Linguistics* 1: 221–41. [CrossRef]
- Landi, Victor. 2005. *The Meyors and Their Language*; Itanagar: Directorate of Research, Government of Arunachal Pradesh.
- LaPolla, Randy J. 2015. Sino-Tibetan syntax. In *The Oxford Handbook of Chinese Linguistics*. Edited by William S.-Y. Wang and Chaofen Sun. Oxford: Oxford University Press, pp. 45–57.
- Lee, Sung-Woo. 2011. Eastern Tamang Grammar Sketch. Master's thesis, Graduate Institute of Applied Linguistics, Dallas, TX, USA.
- Li, Charles, and Sandra Thompson. 1974. An Explanation of Word Order Change SVO => SOV. *Foundations of Language* 12: 201–14.
- Li, Rong, ed. 2002. *Xiàndài hànyǔ fāngyán dà cídiǎn [The Great Dictionary of Modern Chinese Dialects]*. Nanjing: Jiangsu Education Publishing House.
- Li, Xuping. 2018. *A Grammar of Gan Chinese: The Yichun Language*. Berlin: Mouton de Gruyter.
- Li, Yunbing. 2008. Yǔyánjiēchù duì yíxiē nánfāng mínzúyǔyán yǔxù de yǐngxiǎng [The influence of language contact on word order of some minority languages in Southern China]. *Mínzú yǔwén [Minority Languages of China]* 5: 17–34.
- Lieberherr, Ismael. 2017. A Grammar of Bulu Puroik. Doctoral dissertation, Universität Bern, Bern, Switzerland.
- Lieberman, Philip. 2007. The evolution of human speech: Its anatomical and neural bases. *Current Anthropology* 48: 39–66. [CrossRef]
- Lightfoot, David W. 1982. *The Language Lottery: Towards a Biology of Grammars*. Cambridge: MIT Press.
- Lightfoot, David W. 1999. *The Development of Language: Acquisition, Change and Evolution*. Oxford: Wiley-Blackwell.
- Lu, Wen. 2018. Aspects of the Grammar of Tunxi Hui: A Transitional Sinitic Language. Doctoral dissertation, The University of Hong Kong, Hong Kong, China.
- Lü, Shanshan. 2022. *A Reference Grammar of Caijia: An Unclassified Language of Guizhou*. Berlin: Mouton de Gruyter.
- Lupyan, Gary, and Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5: e8559. [CrossRef]



- Manson, Ken. 2009. Prolegomena to reconstructing Proto-Karen. *La Trobe University Working Papers in Linguistics* 12: 1–26.
- Matisoff, James A. 2000. On the uselessness of glottochronology for the subgrouping of Tibeto-Burman. In *Time Depth in Historical Linguistics*. Edited by Colin Renfrew, April McMahon and Larry Trask. Cambridge: McDonald Institute for Archaeological Research, pp. 333–71.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. Berkeley: University of California Press.
- Matthews, Stephen. 1999. Y.R. Chao and Universal Chinese Grammar. In *History of Linguistics 1996 (Vol. 1): Traditions in Linguistics Worldwide*. Edited by David Cram, Andrew R. Linn and Elke Nowak. Amsterdam: John Benjamins, pp. 217–24.
- Matthews, Stephen. 2007. Cantonese grammar in areal perspective. In *Grammars in Contact: A Cross-Linguistic Typology*. Edited by Alexandra Y. Aikhenvald and R. M. W. Dixon. Oxford: Oxford University Press, pp. 220–36.
- Matthews, Stephen, and Virginia Yip. 2011. *Cantonese: A Comprehensive Grammar*, 2nd ed. London: Routledge.
- Miestamo, Matti. 2005. *Standard Negation: The Negation of Declarative Verbal Main Clauses in a Typological Perspective*. Berlin: Mouton de Gruyter.
- Miestamo, Matti, Dik Bakker, and Antti Arppe. 2016. Sampling for variety. *Linguistic Typology* 20: 233–96. [\[CrossRef\]](#)
- Miestamo, Matti. 2013. Symmetric and Asymmetric Standard Negation. In *The World Atlas of Language Structures Online*. Edited by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Nagaraja, K. S. 2010. *Konyak Grammar*. Mysore: Central Institute of Indian Languages.
- Naw, Hsa Eh Ywar. 2013. A Grammar of Kayan Lahta. Master's thesis, Payap University, Chiang Mai, Thailand.
- Naw, Hsar Shee. 2008. A Descriptive Grammar of Geba Karen. Master's thesis, Payap University, Chiang Mai, Thailand.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: The University of Chicago Press.
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Oller, D. Kimbrough. 2000. *The Emergence of the Speech Capacity*. New York: Psychology Press.
- Pan, Wuyun. 1991. Introduction to the Wu dialects. In *Languages and Dialects of China*. (Journal of Chinese Linguistics, Monograph Series 3). Hong Kong: The Chinese University of Hong Kong Press, pp. 237–93.
- Paternicò, Luisa M., and Giorgio Francesco Arcodia. 2023. A synchronic and diachronic analysis of potential *dāk* 得 in Cantonese. *Folia Linguistica*. [\[CrossRef\]](#)
- Peiros, Ilia. 1998. *Comparative Linguistics in Southeast Asia*. Canberra: Australian National University.
- Perreault, Charles, and Sarah Mathew. 2012. Dating the origin of language using phonemic diversity. *PLoS ONE* 7: e35289. [\[CrossRef\]](#)
- Peyraube, Alain. 1991. Some remarks on the history of the Chinese classifiers. *Asian Discourse and Grammar* 3: 106–126.
- Peyraube, Alain. 2015. Grammatical change in Sinitic languages and its relation to typology. In *Diversity in Sinitic Languages*. Edited by Hilary M. Chappell. Oxford: Oxford University Press, pp. 53–78.
- Philippova, Nailya. 2021. A Grammar of Amri Karbi. Doctoral dissertation, University of Helsinki, Helsinki, Finland.
- Pintzuk, Susan. 2017. Variationist Approaches to Syntactic Change. In *The Handbook of Historical Linguistics*. Edited by Brian D. Joseph and Richard D. Janda. Oxford: Wiley-Blackwell, pp. 509–28.
- Plaisier, Helen. 2006. A Grammar of Lepcha. Doctoral dissertation, Universiteit Leiden, Leiden, The Netherlands.
- Post, Mark W. 2017. *The Tangam Language: Grammar, Dictionary and Texts*. Leiden: Brill.
- Rashel, M. Mostafa. 2009. Morphosyntactic Analysis of Mro Language. *Dhaka University Journal of Linguistics* 2: 141–60. [\[CrossRef\]](#)
- Rastogi, Kavita. 2012. *A descriptive grammar of Raji*. Delhi: Aviram Prakashan.
- Rodríguez-Ordóñez, Itxaso. 2019. The role of linguistic ideologies in language contact situations. *Language and Linguistics Compass* 13: e12351. [\[CrossRef\]](#)
- Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill, and Johann-Mattis List. 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences* 116: 10317–22. [\[CrossRef\]](#)
- Sandman, Erika, and Camille Simon. 2016. Tibetan as a “model language” in the Amdo Sprachbund: Evidence from Salar and Wutun. *Journal of South Asian Languages and Linguistics* 3: 85–122. [\[CrossRef\]](#)
- Santorini, Beatrice. 1993. The rate of phrase structure change in the history of Yiddish. *Language Variation and Change* 5: 257–83. [\[CrossRef\]](#)
- Sinnemäki, Kaius, and Francesca Di Garbo. 2018. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology* 9: 1141. [\[CrossRef\]](#)
- Sinnemäki, Kaius. 2020. Linguistic system and sociolinguistic environment as competing factors in linguistic variation: A typological approach. *Journal of Historical Sociolinguistics* 6: 20191010. [\[CrossRef\]](#)
- So-Hartmann, Helga. 2009. *A Descriptive Grammar of Daai Chin*. Berkeley, CA: University of California, Berkeley.
- Stassen, Leon. 2009. *Predicative Possession*. Oxford: Oxford University Press.
- Sun, Chaofen. 1996. *Word-Order Change and Grammaticalization in the History of Chinese*. Stanford: Stanford University Press.
- Sun, Hongkai, and Guangkun Liu. 2009. *A Grammar of Anong. Language Death Under Intense Contact*. Leiden: Brill.
- Szeto, Pui Yiu, and Chingduang Yurayong. 2021. Sinitic as a typological sandwich: Revisiting the notions of Altaicization and Tai-cization. *Linguistic Typology* 25: 551–99. [\[CrossRef\]](#)
- Szeto, Pui Yiu, and Chingduang Yurayong. 2022. Establishing a Sprachbund in the Western Lingnan region: Conceptual and methodological issues. *Folia Linguistica* 56: 25–55. [\[CrossRef\]](#)

- Szeto, Pui Yiu. 2019. Typological Variation across Sinitic Languages: Contact and Convergence. Doctoral dissertation, The University of Hong Kong, Hong Kong, China.
- Szeto, Pui Yiu, Umberto Ansaldi, and Stephen Matthews. 2018. Typological variation across Mandarin dialects: An areal perspective with a quantitative approach. *Linguistic Typology* 22: 233–75. [\[CrossRef\]](#)
- Thomason, Sarah Grey, and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Thomason, Sarah Grey. 2010. Contact Explanations in Linguistics. In *The Handbook of Language Contact*. Edited by Raymond Hickey. Oxford: Wiley-Blackwell, pp. 31–47.
- Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.
- Trudgill, Peter. 2020. Sociolinguistic typology and the speed of linguistic change. *Journal of Historical Sociolinguistics* 6: 20190015. [\[CrossRef\]](#)
- van Driem, George. 2007. The Diversity of the Tibeto-Burman Language Family and the Linguistic Ancestry of Chinese. *Bulletin of Chinese Linguistics* 1: 211–70. [\[CrossRef\]](#)
- Wang, Feng. 2005. On the genetic position of the Bai language. *Cahiers de Linguistique Asie Orientale* 34: 101–27. [\[CrossRef\]](#)
- Wang, Jun. 1991. Language interaction in China. In *Languages and Dialects of China*. (Journal of Chinese Linguistics, Monograph Series 3). Hong Kong: The Chinese University of Hong Kong Press, pp. 161–86.
- Wang, Wenqing. 2007. *Jīnyuán fāngyán yánjiū [Research on Jin Dialect]*. Beijing: Chinese Language Publishing House.
- Weinreich, Uriel. 1974. *Language Contact: Findings and Problems*. 8th printing. The Hague: Mouton de Gruyter.
- Wright, Jonathan Michael. 2009. Hkongso Grammar Sketch. Master's thesis, Graduate Institute of Applied Linguistics, Dallas, TX, USA.
- Wright, Sewall. 1921. Correlation and causation. *Journal of Agricultural Research* 20: 557–85.
- Wu, Yunji. 2005. *A Synchronic and Diachronic Study of the Grammar of the Chinese Xiang Dialects*. Berlin: Mouton de Gruyter.
- Xie, Yan, John MacKinnon, and Dianmo Li. 2004. Study on biogeographical divisions of China. *Biodiversity & Conservation* 13: 1391–417.
- Xu, Dan. 2006. *Typological Change in Chinese Syntax*. Oxford: Oxford University Press.
- Yuan, Jiahua. 1983. *Hànyǔ Fāngyán Gàiyào [An Outline of Chinese Dialects]*, 2nd ed. Beijing: Wenzi Gaige Chubanshe.
- Yue-Hashimoto, Anne O. 1999. *Comparative Chinese Dialectal Grammar: Handbook for Investigators*. Paris: École des Hautes Études en Sciences Sociales.
- Yue-Hashimoto, Anne O. 2003. Chinese dialects: Grammar. In *The Sino-Tibetan Languages*. Edited by Graham Thurgood and Randy J. LaPolla. London: Routledge, pp. 84–125.
- Yue-Hashimoto, Anne O. 2005. *The Dancun Dialect of Taishan*. Hong Kong: City University of Hong Kong.
- Zhang, Menghan, Shi Yan, Wuyun Pan, and Li Jin. 2019. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569: 112–15. [\[CrossRef\]](#)
- Zhou, Chenlei. 2022. *Zhoutun*. London: Routledge.
- Zhou, Chenlei. 2023. Comparative constructions in Zhoutun from a language contact perspective. *Languages* 8: 66. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.