

This is the author-created version of the following work:

White, Nathan (2021) *Grammaticalization and phonological reidentification in White Hmong*. *Studies in Language*, 46 (1) pp. 258-284.

Access to this file is available from:

<https://researchonline.jcu.edu.au/69317/>

Published Version: © John Benjamins Publishing Company. Accepted Version may be made open access under in an Institutional Repository without embargo.

Please refer to the original source for the final version of this work:

<https://doi.org/10.1075/sl.19052.whi>

GRAMMATICALIZATION AND PHONOLOGICAL REIDENTIFICATION IN WHITE
HMONG

Nathan M. White
James Cook University
nathan.white1@my.jcu.edu.au

Nathan M. White
Language and Culture Research Centre
Building D3, James Cook University
14-88 McGregor Rd
Smithfield QLD 4878
Australia

GRAMMATICALIZATION AND PHONOLOGICAL REIDENTIFICATION IN WHITE HMONG¹

Nathan M. White
James Cook University

Abstract

The “dynamic coevolution of meaning and form” of Bybee, Perkins & Pagliuca (1994: 20) has been the subject of significant discussion as regards the languages of Mainland Southeast Asia. However, little work has focused on the mechanisms through which this coevolution occurs when it does surface in these languages.

The current work considers phonological reidentification resulting from phonetic reduction in White Hmong (Hmong-Mien, Laos) involving four morphemes, *ntshai/ntshe* ‘maybe’, *saib/seb* ‘see if/whether; COMP.CFACT’, *puag/pug* ‘LOCL;INTS’, and *niaj/nej* ‘each, every’. These morphemes exhibit an alternation where a rime is phonologically reidentified in a manner consistent with typical phonetic underarticulation patterns, such that an exemplar-model approach (Pierrehumbert 2001, *inter alia*) provides a straightforward explanation.

Furthermore, the data show that the phonological reidentification patterns found in White Hmong exhibit parallels in other languages in the region, confirming that an areal approach to grammaticalization provides greater descriptive adequacy cross-linguistically as regards this phenomenon.

Keywords: grammaticalization; phonological reanalysis; coevolution of meaning and form; verbal modification; linguistic area; Hmong; Southeast Asia

Bybee, Perkins & Pagliuca’s (1994) hypothesis concerning “coevolution of meaning and form” in grammaticalization and how it applies to Mainland Southeast Asian languages has been a focus of discussion in recent years. However, relatively little research has been devoted to identifying the mechanisms through which the phenomenon takes shape in these languages when change in form occurs, or how these mechanisms relate to the fuller cross-linguistic picture.

An exemplar-model approach (Pierrehumbert 2001, *inter alia*) provides a promising avenue for research. The current work explores the phenomenon of changes in form affecting grammatical morphemes in White Hmong (Hmong-Mien, Laos and diaspora) which take the

¹ I would like to thank Sasha Aikhenvald, Nerida Jarkey, and everyone who provided input at SEALS 27, as well as two anonymous reviewers, for their comments and recommendations, some of which have been incorporated into the current work. I would also like to thank Michael Johnson for making available to me some of his comparative data and sources. Any errors or omissions are my own.

form of phonemic reassignment of segments, termed here as “phonological reidentification”. As it will be seen, an exemplar-model approach provides a straightforward explanation of a process of reidentification for the White Hmong data. Furthermore, initial evidence from other tonal Mainland Southeast Asian languages suggests that this process of reidentification is not unique to White Hmong but may apply to a broader linguistic area, and that the exemplar-model approach adequately characterizes the cross-linguistically unusual pattern of form change in languages of the region.

1. Background

Bybee, Perkins & Pagliuca (1994: 115ff.), in their consideration of measuring degrees of grammaticalization, adopt the traditional categories of “isolating”, “agglutinating”, and “fusional” or “inflectional” language.² “Isolating” in their parlance refers to a language that possesses words that are “largely monomorphemic”. Applying their conception of the “dynamic coevolution of meaning and form” (1994: 20) to these three typological categories, they adopt the prediction that these language types should correlate with the degree of semantic development in grammaticalization in a given language. In particular, they claim that “isolating” languages should have the most limited degree of development semantically in terms of abstraction, reflecting the degree of morphological development. Using perfect aspectual forms as a test case, they predict that isolating languages would have perfect-type markers with anterior or completive functions, in contrast to inflectional languages, which would exhibit more

² I place these terms in quotes on the one hand due to the somewhat specific nature of how the terms are used by these authors, and on the other hand due to my reservations about these being typologically motivated categories. These reservations are due to the existence of affixes in many languages commonly labeled isolating in Southeast Asia (q.v. Li & Thompson 1981 and Packard 2000 for Mandarin Chinese, Baxter & Sagart 1997 for Old Chinese, Iwasaki & Ingkaphirom 2005 for Thai, and Thompson 1987 for Vietnamese), which are readily identifiable once a robust, empirically motivated category of “word” is adopted. This applies for Hmong as well (White 2020), though this is outside the scope of the current work.

grammaticalized past or perfective meanings. Using abstract scores derived from their interpretation of data from individual languages, they regard their prediction as confirmed.

Alongside this, Bisang (1996, 2008, 2011, 2015, *inter alia*) proposes and develops a distinctive areal typology for grammaticalization in mainland East and Southeast Asia. In this areal typology, languages possess a high degree of pragmatic inference owing to the complex nature of polysemy with grammatical morphemes—where a single morpheme can have one or more lexical and one or more grammatical functions depending on its distribution and usage. Moreover, the languages exhibit a relative absence of obligatoriness of grammatical items. In Bisang (2008), in particular, he notes that the phonological nature of these languages (in terms of phonotactic restraints and discreteness of syllable boundaries) and multifunctionality of morphemes contribute to a relative lack of Bybee, Perkins & Pagliuca’s coevolution. Instead, semantic development takes the form of widespread pragmatic inference leading to networks of functions, including generalized or abstract ones, for individual morphemes, even where morphological reduction is lacking. Bisang (2015) takes this a step further, linking these observations to the primary vs. secondary grammaticalization distinction made in works such as Kuryłowicz (1965) and Givón (1991). He demonstrates with evidence from several East and mainland Southeast Asian (EMSEA) languages that four criteria for grammaticalization found in the literature—namely, coevolution of meaning and form, constructionalization cycles (drawing on Smirnova 2015), expansion of host class (Himmelmann 2004), and subjectification (Traugott 2010)—do not produce consistent outcomes regarding the distributions and functionality of the morphemes he considers.

A specifically phonological characterization of grammaticalization in EMSEA languages appears in Bisang (2011), where he characterizes phonological change in terms of what can

change—namely, vowel quality, vowel duration, and tonality—and what cannot—syllabicity. As Bisang (2011: 115) describes it, this leads to a scenario where the phonological realization has provided a “degree of resistance” against Bybee, Perkins & Pagliuca’s (1994) coevolution idea. As a manifestation of this “degree of resistance”, Ansaldo, Bisang & Szeto (2018) point out that many grammatical morphemes retain their full phonological content as compared to their original lexical source.

Contributing to the long-term discussion on EMSEA phonological reduction patterns, Ansaldo & Lim (2004) describe phonetic reduction in vowels of grammatical morphemes in Cantonese and Hokkien as compared with their lexical sources. They argue for the existence of reduction in terms of effects on phonological segments without any effect on tone, where the lack of tonal effect is due to the existence of tonal contrast in different registers and a lack of a meaningful neutral tone target. This is contrasted with Mandarin, where tones are often deleted in the course of grammaticalization due to the possibility of an unmarked option in the tone system. For example, Cantonese *gwo*³³ ‘pass, cross; SURPASS’ has a vowel that exhibits centralization, shorter duration, and lower intensity in its grammaticalized form as compared to its lexical form, but the tone is maintained, whereas Mandarin *guo*⁵¹ ‘pass’ undergoes tonal deletion as grammaticalized *guo* ‘EXPERIENTIAL’. Thus, phonological reduction as a component of grammaticalization can be realized through effects on duration and vowel quality even in cases where tones cannot be meaningfully reduced.

Significant to the current discussion is an exemplar-model approach to phonological change. This approach interfaces well with the usage- and frequency-based tendencies of sound change as developed by Bybee (2001, 2011, *inter alia*), and in the past two decades has been the subject of a number of articles and chapters, notably Johnson (1997, 2007), Pierrehumbert (2001,

2002), Wedel (2007), Bybee & Torres (2008), Ernestus & Baayen (2011), and Kirchner (2012), as well as a special issue of *The Linguistic Review* (Gahl & Yu 2006). In this approach, speakers retain phonetic detail of individual tokens of phonemes or words, and individual tokens represent individual “exemplars” stored in the mind of the speaker and used to identify the phoneme or word when listening to output.

In particular, Pierrehumbert (2001) provides a helpful exemplar-model framework for the current discussion, in that she 1) extends the model to cover both identification and production, 2) considers historical change involving reduction and deletion as systematic phonetic bias in connection with hypo-articulation, and 3) incorporates patterns of “entrenchment”. Hypo-articulation in this sense is articulatory undershoot: speakers articulate the sound in a manner that is faster and requires less effort than the manner normally required by the phoneme’s articulatory targets (2001: 147). This then leads to a phonetic realization that is slightly different from what would be expected elsewhere.

Given lenition (in Pierrehumbert’s terminology) and the systematic bias it introduces, a phonetic shift is effectively modeled in that the exemplars are increasingly concentrated in the direction of the hypo-articulation. At the same time, “entrenchment”, i.e., a decrease in variation of a certain form as a result of averaging of activated exemplars in production, models an increase in concentration and strength of certain portions of the distribution of exemplars. She treats phonological neutralization as the final state of a long-term lenition process. This ultimate loss of a distinction emerges as the more-marked form lenites in the direction of a less-marked form and encounters phonetic competition with the latter until the distinction is lost, and the former is incorporated into the latter. This conceptualization adequately characterizes the “phonological reidentification” phenomena described for White Hmong below, and thus

Pierrehumbert's (2001) framework is incorporated into my working definitions on phonological change below.

1.1 Working definitions and scope

Considering the various definitions used in the literature as to what represents “mainland Southeast Asia,” it is necessary to adopt a clear definition for the current work. Given its adequacy in roughly delineating the region containing the languages affected by the typological patterns discussed in Bisang (1996, etc.) as described above, I adopt the definition proposed by Enfield (2017: 601): “the area occupied by present-day Cambodia, Laos, Peninsular Malaysia, Thailand, Myanmar and Vietnam, along with areas of China south of the Yangtze River.” This definition enables the inclusion of the entire dialect continuum of Hmong, as well as the other languages in Bisang's work and other sources affected by high pragmatic inference and polysemy.

Furthermore, terminology surrounding “reduction” has led to inconsistencies across the literature, especially in relation to “phonetic” vs. “phonological reduction.” For the current work, “phonetic reduction” is taken as the purely automatic reduction that results from fluency, fast speech, and frequency—that is, a reflection of Pierrehumbert's (2001) modeling of phonetic bias resulting from hypo-articulation. “Phonological reduction”, on the other hand, is adopted as a banner term for those changes that represent a genuine change in phonological understanding or awareness. As it can be seen, “phonetic reduction” is the most relevant to early stages of grammaticalization, where frequent words most often undergo automatic reduction due to a high rate of speech but before the phonemes in a grammatical word begin to change to something else (“phonological reduction”). More specifically, a changing from one phoneme in the inventory to another in terms of phonological identification, which falls into the “neutralization” category of Pierrehumbert's (2001) framework, can be termed for the current purposes “phonological

reidentification,” on the grounds that a phoneme comes to be identified as another. Using “reidentification” rather than “neutralization” is preferable as neutralization can cover a much broader range of meaning, including phenomena such as systemic final devoicing (q.v. Pierrehumbert 2001: 152ff.). Phonological reidentification is thus a subtype of phonological reduction, where phonological reduction can affect more than just identification of phonemes in a system—e.g. phonemic loss, merging of independent phonemes, etc.

While phonological reidentification reasonably encompasses both segmental and suprasegmental phenomena, such as tone, much of the existing literature relevant to the exemplar model approach as it pertains to phonology discusses only segmental phenomena. However, tonal phenomena in Hmong and other Hmong-Mien languages (see for example section 2.3 below) suggest that tonal change and reidentification of segments operate along parallel lines. Nevertheless, the current work takes segmental change as its scope, demonstrating how the exemplar-model approach can be applied to explain the phenomenon of phonological reidentification of segments in White Hmong. Tonal change and its relationship to phonological reidentification in grammaticalization are left for further research.

For the current consideration of phonological reidentification in Hmong, several principles must be discussed involving sound correspondences and recognition of grammaticalization. First, the Hmong-Mien languages occasionally exhibit idiosyncratic changes in their sound correspondences, where one language will have a form that differs from the sound correspondence pattern shared by all other languages. As a result, I adopt the following principle given the discussion in White (2021): “if a form has at least three shared categories between onset, rime, tone, and meaning, then it is deemed a cognate, especially if the form of the remaining category is reasonably similar or a common result of reduction.” In the area of

meaning, this can involve grammatical function, as grammaticalization often involves semantic change and bleaching.

Moreover, data in Hmong-Mien languages are generally synchronic: very few, if any, diachronic data are ever available for any one language. As a result, demonstrating the existence of grammaticalization relies on the following principles:

- 1) A given form should have two meanings/functions in the language or a closely related language, where one is more grammatical than the other and so exhibits semantic shift or bleaching characteristic of grammaticalization;
- 2) A construction that can be recognized as a bridging construction should exist in the language or a closely related language; and
- 3) Other features of grammaticalization should correlate with the two forms in principle 1, such as host class expansion or decategorialization.

“Closely related language” featured above reflects the fact that the original grammaticalization could have taken place at the immediate parent language stage. If this is the case, the language under consideration (White Hmong, in this case) may only retain the more grammaticalized form, while the closely related language may contain both the more lexical and more grammaticalized form. Conversely, if the grammaticalization took place after divergence from the parent language, then the language under consideration should have both forms, while the closely related language (and other related languages) will only have the more lexical form. Each of these factors will prove critical in the discussion of *puag/pug* ‘LOCL;INTS’ and *nij/nej* ‘each, every’ below.

1.2 Language

The language considered in the current work is White Hmong (Hmong-Mien, Laos), with special reference to White Hmong as spoken in diaspora communities in Australia and the United States.

White Hmong, in its original context, is spoken in Laos, near the southern end of a dialect

continuum that extends from Thailand in the south to southwestern China in the north. Besides this dialect continuum, the Hmongic branch of the Hmong-Mien family possesses a number of other subbranches (Ratliff 1992; Taguchi 2013; Sposato 2015), as displayed in Figure 1 below.

[Figure 1 Here]

As discussed in White (2020), Hmong has three criteria that define a grammatical word: 1) lexically specific tone alternations, 2) grammatical cohesion in terms of ability to isolate and inability to separate, and 3) reduplication. Given these criteria, both multisyllabic grammatical words and dozens of affixes can be recognized in White Hmong. Nevertheless, the language is “isolating” in Payne’s (2017) sense of having a relatively low score on an index of synthesis as it has relatively few affixes per word. At the same time, White Hmong possesses a range of grammatical morphemes required to co-occur with certain other elements such as classifiers, nouns, or verbs. Phonologically, it has a rich system of segmental phonemes, both consonants and vowels, as well as a system of seven contrastive tones.³ The consonant system appears in Table 1 below, vowels and diphthongs appear in Tables 2 and 3, respectively, and the system of tones is displayed in Table 4. Romanized Popular Alphabet orthographic representation is provided in brackets next to each phoneme and tone in each table.

[Table 1 Here]

[Table 2 Here]

[Table 3 Here]

[Table 4 Here]

One important feature of White Hmong phonology is the system of phonetic reduction patterns attested. In consonants, some phonemes undergo weakening: fricative phonemes, for example, weaken to approximants, as with /v/ and /j/, which are often articulated as [v] and [j],

³ The discussion of the phonological system of White Hmong here closely follows White (2020).

respectively. This weakening is a gradient phonetic process in Hmong, and is a product of speed and relaxed enunciation, that is, a product of hypo-articulation in Pierrehumbert’s (2001) sense.

Vowels, likewise, can undergo a pattern of gradient phonetic reduction in rapid, connected speech. This especially applies to grammatical morphemes—including closed class items like pronouns—and initial syllables in multi-syllabic words, which may or may not be separate morphemes. Diphthongs can reduce to a simple vowel that reflects either the first of its two constituent elements or a coalescing of the two elements; with simple vowels (including those resulting from diphthong reduction), reduction can occur toward mid in height or central in backness, or even deletion for those elements originating as individual vowels. Examples appear in (1-3) below.

- (1) *lawv* *tsis* *tuaj*
 / \widehat{lei}^{24} $\widehat{tʃi}^{22}$ $tʃeɯ^{53} \%L^4$
 [$l\text{ə}^{22}$ $tʃ$ $tʃeɯ^{52}$]
 3PL NEG come
 they did not come

- (2) *no~no*
 / $\text{np}^{33} \text{np}^{33}$ /
 [$\text{np}^{22} \text{np}^{33}$]
 REDUP~be.cold
 be very cold

- (3) *tas* *li* *xwb*
 / $t\text{a}^{22}$ li^{33} ʃi^{55} /
 [$t\text{a}^{22}$ le^{33} ʃi^{55}]
 finish INTS only
 always

In (1), *tsis* ‘NEG’ is reduced from $\widehat{tʃi}^{22}$ to $[tʃ]$, where the vowel /i/ has been deleted, and the diphthong \widehat{ei} in *lawv* ‘3PL’ is phonetically reduced to the mid central [ə]. In (2), the vowel of the reduplicant portion of *no no* ‘be very cold’, phonologically / v /, is realized as a more centralized

⁴ Here, %L represents a low boundary tone in the Hmong intonation system.

[ɣ], while in (3), *li* ‘INTS’, which is phonologically /li³³/, is phonetically realized as [le³³], where the vowel is phonetically realized as mid in height.

Certain rimes in White Hmong may have an additional phonemic property of nasality, which may be termed “phonological rime nasalization.” This phonological nasalization may take the form of a nasal coda consonant, nasalization on the vowel, both, or in rapid speech neither, with no clear determining factors that would serve to otherwise predict the surface form. Its phonemic status is expressed through minimal pairs such as *nej* [ne⁵³] ‘you (pl.)’ and *nej* [nẽŋ⁵³] ‘life’.⁵

Tone in White Hmong also shows signs of simplification in rapid, connected speech. The system of tone reduction takes the form of gradient phonetic hypo-articulation in the contour of tones given neighboring tones. Typically, this involves simplification of a contour to a level pitch. Tones such as <v> /²⁴/ or <j> /⁵³/ may be realized as level tones, alongside the creaky <m> /³¹/ tone, and the breathy <g> /^{42/53}/ tone, which can appear with a reduced fall. Examples of this phenomenon appear below.

- (4) *kuv* *lub* *npe*
 /kø²⁴ lø⁵⁵ m^{pe}³³/
 [kø²² lõ⁵⁵ m^{pe}³³]
 1SG CLF;GENER name
 my name
- (5) [...] *uas* *nws* *tuaj* *kawm* *nyob*
 /ʔø̃²² ɲi²² tø̃⁵³ kə̃³¹ ɲw⁵⁵/
 [ʔø̃²² ɲi²² tø̃⁵⁵ kə̃³² ɲw⁵⁵]
 REL 3SG come study be.at
 (who) came to study in

⁵ Note also that there are some phonological alternations in a small number of lexemes containing the vowel /ø/: for example, *hnub/hnoob* ‘day’, and *nug/noog* ‘ask’. These seem to be the result of idiosyncratic variation rather than frequency effects, as both variants regularly appear regardless of speed. In fact, Xiong & Cohen (2005) attest a rime of <o> not only for *hnub*, but also for *hloov* ‘exchange’, suggesting that this phenomenon is one of dialectal variation rather than phonological reidentification due to reduction.

Here, in (4), the /²⁴/ vowel in *kuv* ‘1SG’ appears as a level [²²], while in (5), the /⁵³/ tone of *tuaj* ‘come’ is realized as a level [⁵⁵] and the /³¹/ creaky tone of *kawm* ‘study’ appears as a reduced [³²] fall.

1.3 Data sources

Sources for the current research include several consultants of Laotian Hmong residing in diaspora locations in Australia and the United States, as well as corpus data from the soc.culture.hmong (SCH) Usenet corpus (Mortensen 2015) as well as the current author’s own Chuanqiandian (CQD) corpus (White n.d.) containing Chinese Hmong data drawn from Chinese-based Hmong social websites.

The corpora contain a wealth of data, though the speaker composing the original content is typically not identifiable, and thus the exact speech variety may not be perfectly identifiable. Note, however, that, in the case of the SCH corpus, this is counterbalanced by the large number of contrastive phonological and lexical content that Laotian White Hmong contains versus other Hmong varieties, as well as the increasing degree of lexical and grammatical convergence displayed between Hmong varieties in the diaspora. In addition, apart from those corpus elements that constitute formal literary productions (songs, stories, etc.), production errors do appear, and thus each SCH corpus example has been checked with native speakers. As for the CQD corpus, due to the lack of readily available speakers of Chinese Hmong dialects, frequency of attestation alone has been used to distinguish common elements from production errors (which naturally would tend to be rarer), while noting that a significant proportion of the CQD corpus content is of a formal literary nature. In any case, the CQD data is used here in a limited role to supplement other sources.

Furthermore, given that many of the examples discussed below, especially those from the corpora, are drawn from written sources, the use of a certain grapheme in what is effectively a

phonemic script are regarded as the writer's identification of a phoneme with other instances of that phoneme in other words.

Data considered for Weining Ahmao are found in Wang (1972, 1986), GMYZWY (1965), and Parsons & Parsons (2001), as well as the text collection found in the Hua Miao Archive (Rake n.d.). Finally, other Western Hmongic data reviewed for potential comparative relationships include material on Hmyo (Luobohe/Luobo River Miao; Taguchi 2008, 2012, n.d.), Mang (Mashan Miao; Wu & Yang 2010), several varieties of Bunu (Meng 2001), and the comparative Hmong-Mien datalist found in Chen (2013).

1.4 Organization

The current work contains four sections: while the first has provided an introduction to the topic, section 2 focuses on several grammatical morphemes associated with phonological reidentification, section 3 ties the phenomena associated with these morphemes to the broader areal and typological perspective, and section 4 provides a brief conclusion summarizing findings.

2. The data

White Hmong has a number of grammatical morphemes containing phonemes affected in their realization by alternations that suggest phonological reidentification. Four morphemes that form the focus of the current inquiry are: *ntshai/ntshe* 'maybe', *saib/seb* 'see if/whether; COMP.CFACT', *puag/pug* 'LOCL;INTS', and *niaj/nej* 'each, every'. Each of these morphemes are considered in turn.

2.1 *ntshai/ntshe* 'maybe'

This first morpheme has its diachronic origins in the concurrent transitive verb *ntshai* 'be afraid of', for example:

- (6) *Lawv_A ntshai [tus dev]_O.*
 3PL be.afraid.of CLF;ANIM dog
 They are afraid of the dog.

As a grammatical operator on the clause preceding the clausal subject, this morpheme varies as *ntshai/ntshe*, where the diphthong / \widehat{ai} / may exhibit reidentification as /e/, and signals a moderate degree of certainty (White 2014), with a possible translation as ‘maybe’ (Xiong 2017), as follows:

- (7) *Ntshai koj_A tsis nyiam.*
 maybe 2SG NEG like
 Maybe you don’t like (it). (Xiong 2017)

- (8) *Ntshe koj_A nyiam.*
 maybe 2SG Like
 Maybe you like (it). (Xiong 2017)

Note that attempting to substitute the variant *ntshe* in (6) above results in an unacceptable sentence, as in (6’) below.⁶

- (6’) **Lawv_A ntshe [tus dev]_O.*
 3PL CLF;ANIM dog
 Intended: They are afraid of the dog.

This demonstrates that the variant form *ntshe* is generally limited to the grammaticalized expression of *ntshai* with the degree of certainty meaning ‘maybe’.

Fueling this alternation are the phonetic reduction patterns that normally affect grammatical morphemes. As discussed in section 1.2 above, a possible realization of phonetic reduction on a vowel can take the form of neutralization of height to mid, and in the case of *ntshai* this specific form of reduction would be realized as [$\widehat{t^h e}$]. One would expect this pronunciation to be written as *ntshe* if the pronunciation were in fact interpreted as a straightforward reflection of the underlying phonemes, where [e] is interpreted as /e/ rather than

⁶ It should be noted here that a form occurring with verbal morphology *ntshe* does appear, though quite rarely, in Mortensen (2015); nevertheless, White Hmong speakers consulted generally reject *ntshe* as a verb.

as a reduction from /*aî*/. This suggests that phonological reidentification is involved in that a rapid pronunciation with [e]—a relatively common one for a grammatical morpheme such as this—has come to be taken by some speakers as either the basic phonemic representation as /e/ or as taking part in a free alternation of variant forms containing either /e/ or /*aî*/.⁷

2.2 *saib/seb* ‘see if/whether; COMP.CFACT’

The morpheme *saib/seb* has its origins in the verb *saib* ‘look at’:

- (9) *Ces* [*cov caub.nab.thib ntawd*]_{S/A} *tuaj saib* [*tus neeg*
 then CLF;PA official LOCL;nearby come look.at CLF;ANIM person

tuag_{RC} no]_O *tas...*
 die this finish
 Then after those officials came to look at the dead person...

In its function as a grammatical entity, this morpheme is realized as *saib* or *seb* and exhibits two degrees of grammaticalization: 1) it serves in a grammatically intermediate role between verb and complementizer to supply the meaning ‘see if/whether’ (as in (10)), and 2) serves in a further grammaticalized role derived from the first to introduce a complement clause as ‘COMP.CFACT’ (as in (11)).

- (10) *Tsuas los nyob tom_E seb yus_S yuav rov qab_E...*
 only come.back stay LOCL;there see.if INDF IRR return back
 We only came to stay there to see whether we would return...
- (11) *Kuv_A xav paub [seb [qhov no]_{CS} yog dab.tsi_{CC}]_{CoCl:O}*
 1SG want know COMP.CFACT CLF;DEFAULT this COP what
 I want to know what this thing is.

Similarly to *ntshai/ntshe* ‘maybe’ above, *saib/seb* ‘see if/whether; COMP.CFACT’ displays a phonological alternation between /*aî*/ and /e/ in its grammaticalized functions. As with *ntshe* ‘maybe’, /*aî*/ can be realized as [e] as a result of neutralization to a monophthong realized with

⁷ Interestingly enough, Xiong & Cohen (2005) only attest a form <*nchait*>, a perfect phonological match for *ntshai*, in this function; likewise, the CQD corpus only contains instances of <*nchait*>—an equivalent for *ntshe* never appears. This evidence thus reinforces the interpretation of the data found here.

mid height. The morpheme as *saib* would then be realized as [ʃe⁵⁵], which, if taken as phonologically transparent, would be reinterpreted as *seb* /ʃe⁵⁵/, which is the form typically observed. Note that the change to *seb* appears with both degrees of grammaticalization, not merely the second.

2.3 *puag/pug* ‘LOCL;INTS’

This morpheme is a localizer (term following Xiong & Cohen 2005: 63ff.; cf. “spatial deictic” in Ratliff 1990 and Jarkey 2015) that typically appears with other localizers or with demonstratives to intensify the deictic value of that other element. An example appears in (12).

- (12) ...*kuv*_S *ya* [*puag* *saum* *ib-nta* *ntuj*]_{PH} *xwb*.
 1SG fly LOCL;INTS LOCL;on.top.of one-CLF;MIDPOINT sky only
 I will just fly high up in the sky. (Johnson & Yang 1992: 325)

Above, *puag* ‘LOCL;INTS’ appears with another localizer, *saum* ‘LOCL;on top of’, to indicate that the location of the event would be far from the addressee.

Diachronically, Weining Ahmao, a Far Western Hmongic language closely related to Hmong, provides evidence of a bridging construction where *bhi*^{35/31} ‘DEM;MED.HIGH’⁸ (cognate with WH *pem* ‘LOCL;ABOVE’) is used to modify another deictic word to intensify the distance associated with deictic reference. The White Hmong form *puag* ‘LOCL;INTS’ shares onset, tone, and this function with the Weining Ahmao form, thus fulfilling the principle of cognacy presented in §1.1 above; the rime *ua* /øǣ/ is innovative, perhaps the result of an earlier idiosyncratic shift from /e/ to /a/ (which systematically led to /øǣ/ in White Hmong). The function of *puag* ‘LOCL;INTS’ is reflective of the bridging construction meaning in Weining Ahmao, exhibiting semantic shift from the value represented by the modern White Hmong *pem* ‘LOCL;ABOVE’ in other contexts; both values are represented in the distribution of Weining Ahmao *bif* ‘DEM;MED.HIGH’. *Puag* ‘LOCL;INTS’ shows a much wider type distribution in White

⁸ Gloss and transcription following Gerner (2009: 80-82).

Hmong as compared to Weining Ahmao, thus exhibiting host class expansion. Thus, a source at the Proto-Far Western Hmongic stage shared with *pem* ‘LOCL;ABOVE’ would be justified given the principles presented in §1.1 above. To be sure, one could argue that these forms have converged in Weining Ahmao, and White Hmong retains the two distinct forms; however, other Western Hmongic varieties reviewed only attest cognates for WH *pem* ‘LOCL;ABOVE’, with no independent trace of *puag* ‘LOCL;INTS’ in form or function. Thus, I adopt the analysis above as the origin of *puag* ‘LOCL;INTS’ for purposes of the current discussion.

As a development in White Hmong, some speakers today consistently write *puag* ‘LOCL;INTS’ as a monophthongal form *pug*, reflecting their interpretation of the spoken form, as in the following example.

- (13) [*Dav*_{S/A} *yuav los tom qaib*_O *mas*]_{TOP} *nws*_{S/S/S/A} *los*
hawk IRR come bite chicken TOP 3SG come
- mus zaum [pug tim tej ceg ntoo]*_{PH} *saib*
go sit LOCL;INTS LOCL;on.other.side CLF;SOME branch tree watch
- [*cov qaib*]_O.
CLF;PA chicken
- When a (specific) hawk would attempt to catch the chicken, it would come to sit over on some tree branches and watch the chickens. (Cha 1994: 26)

In this case, *puag* ‘LOCL;INTS’ appears as *pug* in the phrase *pug tim tej ceg ntoo* ‘over on some tree branches’. Note that Cha (1994) contains dozens of examples of *pug* ‘LOCL;INTS’, such as the one above, providing evidence for the conventionalization of this change.

The variation between *puag* and *pug* follows along the lines of the phonetic realization of *tuaj* (/tʰəŋ⁵³/) ‘come’ as [tʰə⁵⁵] in (5) above: the earlier form is *puag*, where the diphthong /əŋ/ is phonetically realized as the monophthong [ə] due to hypo-articulation (q.v. §1.1). This is then ostensibly conventionalized by some speakers and reidentified as /ə/, taking the monophthongized form *pug*.

2.4 *niaj/nej* ‘each, every’

The word *niaj* ‘each, every’ is a quantifier that is generally realized as *niaj*, as in the following example.

- (14) *txog rau niaj hnub no*
 until to every CLF;DAY this
 until (every one of) these days (Cha 1994: 176)

Diachronically, the evidence suggests that this word originates in *niaj*, a borrowing from Chinese 年 (SMC *nián*) ‘year’, a term used elsewhere in Hmong in expressions for years. Comparative data from Weining Ahmao and other Hmong varieties as well as internal evidence from White Hmong suggest that this term originally meant ‘year’. It was used with the native Hmong *xyoo* ‘year’ to form a coordinate compound *niaj + xyoo*, attested in a Chinese Hmong variety as *<nax xongt>* ‘year, age, year’s harvest’ (GMYZWY & ZKYSMYDDEGD 1958: 221) and in Weining Ahmao as *<niex xiaot>* ‘age, year’s harvest’ (GMYZWY 1965: 217), as well as to serve in parallel in ABCB patterns in four-part elaborate expressions. This coordinate compound construction and elaborate expressions such as that found in (15) below, provided the bridging construction context of *niaj* as involving ‘each’ rather than ‘year’.

- (15)
- | | | | | |
|-----------|---------------------|--------------------|---------------------|---------------------------------|
| | <i>txhua</i> | <u><i>niaj</i></u> | <u><i>txhua</i></u> | <u><i>xyoo</i></u> ⁹ |
| Original: | [each] | [year] | [each] | [year] |
| Reparsed: | [| each |] | [year] |
| | each and every year | | | |

In (15), *niaj* historically (evidenced by the exact cognate expression in Weining Ahmao) was interpreted such that it represented ‘year’ in parallel to *xyoo* ‘year’, while in recent use, *txhua_niaj_txhua* can be interpreted as a chunk supplying the meaning ‘each’, which has been applied to use with *zaus* ‘CLF;TIME’ to mean ‘each time’. This semantic shift aligns with evidence of host class expansion: *niaj* ‘each’ is used with *xyoo* ‘year’ as well as several other classifiers

⁹ The use of the underscore in this example follows White (2020): elaborate expressions often contain elements that together exhibit an intermediate wordhood status (indicated using underscores), as is the case here.

signaling time reference, and occasionally with nouns such as *Hmoob* ‘Hmong’. As with *puag* ‘LOCL;INTS’, one could argue an alternative analysis: *niaj* ‘each, every’ is historically a separate, unrelated entity from *niaj* ‘year’. However, this fails to explain the strong relationship between *niaj* ‘each, every’ and time reference, nor does it explain the fact that Western Hmongic languages reviewed lack a cognate for *niaj* ‘each, every’ while Weining Ahmao attests a cognate for *niaj* ‘year’ that does not exhibit evidence of a separate borrowing from Chinese. As a result, for purposes of the current discussion, I adopt the analysis as presented above.

Synchronically, based on his work with White Hmong speakers in Thailand, Mottin (1978: 55) reports variation between *niaj* and *nej*. Two examples from his work appear below, the first with *niaj* (in (16)) and the second with *nej* (in (17)).

(16) *niaj tus neeg*
 each CLF;ANIM human.being
 each person (Mottin 1978: 55)

(17) [*Nej hnuv*]_{PH:TOP} *yus_A ua hauj.lwm_O tas zog_O*.
 each CLF;DAY INDF do work finish strength
 Every day, I do my best. (Mottin 1978: 44)

It should be noted here that the available data for Thai and Laotian White Hmong strongly suggest that these two differ primarily by political geography rather than by another measure. If the same phonetic hypo-articulation patterns (q.v. §1.1) described above apply in this case, then the alternation is explained: the /iɤ/ diphthong is subject to hypo-articulation as a mid vowel [e], which is then reidentified as /e/. This phenomenon reflects a phonological change toward monophthongization affecting a morpheme that likely has undergone grammaticalization from a lexeme at an earlier stage.

3. Discussion

Each of the morphemes above represent (likely) grammaticalized entities that each have undergone phonological reidentification as a result of phonetic hypo-articulation. First, *ntshai*

has its origins as a verb meaning ‘be afraid of’, compared to a grammatical counterpart *ntshai/ntshe* as a degree of certainty marker meaning ‘maybe’, where the /*ai̯*/ diphthong of the original lexical item has been reidentified as /*e*/. This likewise applies with the verb *saib* ‘look at’, which has grammaticalized to become *saib/seb* ‘see if/whether; COMP.CFACT’, following the same phonetic trajectory that has led to reidentification of the vowel as with *ntshai/ntshe* ‘maybe’. *Puag* ‘LOCL;INTS’ appears alternately in the usage of some speakers as *pug*, exhibiting a shift from /*œ̯*/ to /*ø*/, after possibly having undergone a prior reanalysis in early Hmong. Finally, *niaj* ‘each, every’ has likely undergone a highly idiosyncratic grammaticalization process from ‘year’ to ‘each, every’, and has been reported as an alternative *nej* in a Thai White Hmong lect. This is fully consistent with phonetic hypo-articulation of a diphthong as a monophthong representing a coalescence of the elements commonly observed in White Hmong, from /*i̯ə̯*/ to /*e*/.

In each case, the simpler, monophthongized vowel appears in a grammaticalized form, and the phonological identity of this vowel is fully consistent with observed phonetic hypo-articulation practices among White Hmong speakers of the original vowel. The best explanation for this phenomenon is that the pervasive phonetic reduction found in fluent Hmong speech (as in §1.2 above), which targets especially grammatical morphemes, has led to a long-term phonetic bias effect in terms of Pierrehumbert (2001). This then has resulted in speakers’ reidentification of the diphthong as a monophthong that is phonetically realized the same way as the segment that has experienced the phonetic bias effect.

On the one hand, phonetic reduction in grammatical morphemes is expected cross-linguistically, especially along the lines of coevolution of meaning and form (cf. Bybee, Perkins & Pagliuca 1994). On the other hand, the forms described above all have in common the fact that they do not exhibit morphological reduction (cf. Bisang 2008) nor do they lose their status as

independent syllables (cf. Bisang 2011). Furthermore, in regard to primary vs. secondary grammaticalization as discussed in Kuryłowicz (1965) and Givón (1991), the phonological reidentification exhibited above is manifested with *saib/seb* in its primary grammaticalization from a full verb *saib* ‘look at’ to *saib/seb* ‘(see) if/whether’ with no further change at the secondary grammaticalization stage as ‘COMP.CFACT’. This is comparable to Bisang’s (2015) contention that Mainland Southeast Asian Languages fail to exhibit the primary-secondary grammaticalization distinction in terms of phonological change, again confirming the areal nature of phonological change in grammaticalization.

In addition to the grammatical morphemes described above, initial evidence suggests that other White Hmong morphemes may be affected by phonological reidentification as well. One of these is *no(v)* ‘this’, which historically descends from Proto-Far Western Hmongic **naŋ^{B1}*, where evidence of a historical rime in **aŋ* includes the fact that *no(v)* belongs to a rime class reflecting Proto-Hmongic **ein* (Ratliff 2010: 172) embodied in Mile Hmong *naŋ⁴⁴* ‘this’ (Li 2015: 47ff.), yet contains a rime in /v/. Another is *puas/pis/pes tsawg* ‘how many’, for which the first syllable is attested across Hmong dialects with a range *pəŋ/pə/po/pi/pe* (Lyman 1979, Savina 1916, Wang 1985, Xiong & Cohen 2005, *inter alia*); the dialectal divergence may ultimately prove to be the result of the development of variation in phonetic bias in Pierrehumbert’s (2001) sense between local dialects which then led to different reidentification results.

3.1 Areal and typological perspective

The pattern of phonological reidentification accompanying grammaticalization in White Hmong has parallels in a number of unrelated languages in the larger region. First, it is comparable to a pattern of reidentification in Standard Mandarin Chinese (cf. Bisang 2008: 31). In these cases, complex codas in Mandarin are simplified to [ə] and recognized as /ɤ/ with loss of tone, and feature especially in grammatical morphemes such as *le* ‘COMPLETION’, which historically

originates in the verb *liǎo* ‘bring to completion’ (Yip & Rimmington 2004: 102), or *-zhe* ‘DURATIVE’ which derives from *zhuó* ‘touch, reach’ (Arcodia 2013: 7).¹⁰

Among other Mainland Southeast Asian languages, Vietnamese shows a variation in the verbal modifier *chǎng* ‘definitely not’ as either with a tone change in *chǎng* or a rime simplification as *chả* (Thompson 1987: 210). Thai likewise shows a rime simplification in the verb *càk* ‘desire, intend’ to the irrealis marker *ca* (Diller 2001: 158). This is alongside Diller’s (2001: 166) speculation that *lae*^{A2} ‘then’ and subsequently *lae*^{D2} ‘and’ stem from *lae:w*^{C2} ‘finish’ as does *la*^{D2} ‘also’. Lao (Enfield 2007) exhibits even more robust simplification in vowel length reduction and monophthongization alongside tone deletion in verb-modifying morphemes, modifier classifiers, topic markers, and class term and kinship-based classificatory prefixes when compared to their lexical and grammatical sources.

Each of these tonal languages from Mainland Southeast Asia exhibit rime reduction with or without tonal reduction alongside a general resistance to reduction to a unit smaller than the syllable (cf. Bisang 2011). At the phonetic level, a hypothesis involving phonetic reduction of the rime of grammatical elements with maintenance of syllabicity and tone is confirmed for Cantonese and Hokkien (Ansaldò & Lim 2004), while the data for White Hmong presented above exhibit the conventionalization of phonetic reduction through reanalysis in the form of reidentification at the phonological level. The data from these tonal Mainland Southeast Asian languages strongly suggest the same trajectory is in play: grammaticalization-related reduction may target the rime of the syllable, but rather than allow the reduction of the morpheme to a subsyllabic element, the rime instead may be phonologically reidentified in these languages as containing simpler phonemic content. For these languages, this is the primary manifestation of Bybee’s coevolution of meaning and form where it does occur at all.

¹⁰ Note that Bisang (2011: 115) states that this marker is derived instead from *zhù* ‘live, reside, stay’.

The current perspective of rime reduction and phonological reidentification in the course of grammaticalization is reminiscent of Heine & Kuteva's (2005, *inter alia*) idea of a grammaticalization area, that is, a geographic area where languages share common grammaticalization patterns that derive from contact. Taking this conceptualization in light of Narrog & Heine's (2018) perspective of phonetic erosion as the feature of grammaticalization most sensitive to typological variation, the cross-linguistically uncommon reduction/reidentification patterns shared between White Hmong and these other languages can be regarded as an areally specific phenomenon, contributing to the larger picture of a grammaticalization area.

What makes this phenomenon typologically interesting, however, is that this form of phonological reidentification is a reasonably common part of the grammaticalization process in White Hmong, yet the current perspective regarding grammaticalization areas in the literature tends to focus on shared grammatical innovations derived from contact, rather than phonological ones. Here, a phonological process specific to grammatical and grammaticalizing morphemes is shared in a historical contact situation. This process is likely the result of both contact—in the form of influence on the phonological structure of word and syllable—and the effects of typical diachronic internal pressure—phonetic reduction associated with the general process of grammaticalization—which has then led to a degree of areal convergence.

4. Conclusion

Altogether, White Hmong has a set of grammatical morphemes that exhibit an optional phonological reidentification of the rime, where a more complex rime is reidentified as a simpler one with the syllabicity of the morphemes retained. The exemplar-model approach as presented in Pierrehumbert (2001) provides a straightforward mechanism for this process, where phonetic

hypo-articulation leads to the accumulation of a phonetic bias, which, in the case of White Hmong, leads to reanalysis as a different phoneme. Initial evidence suggests that reduction and reanalysis of the rime of grammatical and grammaticalizing morphemes combined with maintenance of the syllable is shared with other Mainland Southeast Asian languages. This is consistent with prior findings regarding the areal nature of how changes in form are manifested. The grammaticalization area concept of Heine & Kuteva (2005, *inter alia*), adequately characterizes this phenomenon for languages in the region, which ultimately reflects the areal nature of major cross-linguistic differences in grammaticalization.

References

- Ansaldo, Umberto & Lisa Lim. 2004. Phonetic absence as syntactic prominence: Grammaticalization in isolating tonal languages. In Fischer, Olga, Muriel Norde & Harry Perridon (eds.), *Up and down the cline—The nature of grammaticalization*. Amsterdam: John Benjamins Publishing Company.
- Ansaldo, Umberto, Walter Bisang & Pui Yiu Szeto. 2018. Grammaticalization in isolating languages and the notion of complexity. In Narrog, Heiko & Bernd Heine (eds.), *Grammaticalization from a typological perspective*, 219-234. Oxford: Oxford University Press.
- Arcodia, Giorgio Francesco. 2013. Grammaticalisation with coevolution of form and meaning in East Asia? Evidence from Sinitic. *Language Sciences* 40:148-167.
- Baxter, William H. and Laurent Sagart. 1997. Morphology, Phonology, and the Lexicon in Modern and Ancient Chinese. In Packard, Jerome L. (ed.), *New Approaches to Chinese Word Formation*. Berlin: Mouton de Gruyter.

- Bisang, Walter. 1996. Areal typology and grammaticalization: Processes of grammaticalization based on nouns and verbs in East and mainland South East Asian languages. *Studies in Language* 20:3, 519-597.
- Bisang, Walter. 2008. Grammaticalization and the areal factor—the perspective of East and mainland Southeast Asian languages. In López-Couso, María José & Elena Seoane (eds.), *Rethinking grammaticalization: New perspectives*. Amsterdam: John Benjamins Publishing Company.
- Bisang, Walter. 2011. Grammaticalization and linguistic typology. In Narrog, Heiko & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 105-117. Oxford: Oxford University Press.
- Bisang, Walter. 2015. Problems with primary vs. secondary grammaticalization: The case of East and mainland Southeast Asian languages. *Language Sciences* 47, 132-147.
- Bybee, Joan & Rena Torres Cacoullos. 2008. Phonological and grammatical variation in exemplar models. *Studies in Hispanic and Lusophone Linguistics* 1:2, 399-414.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, Joan. 2011. Usage-based theory and grammaticalization. In Narrog, Heiko and Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 69-78. Oxford: Oxford University Press.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press.
- Cha, Tua. 1994. *Kawm ntawv Hmoob*. Fresno, USA: Fresno City College.

- Chen, Qiguang. 2013. *Miao-yao yuwen*. [Miao and Yao language]. Beijing: Central University for Nationalities Press.
- Diller, Anthony V.N. 2001. Grammaticalization and Tai syntactic change. In Kalaya Tingsabadh, M.R. and Arthur S. Abramson (eds.), *Essays in Tai linguistics*, 205-228. Bangkok: Chulalongkorn University Press.
- Enfield, N.J. 2007. *A grammar of Lao*. Berlin: Mouton de Gruyter.
- Enfield, N.J. 2017. Language in the mainland southeast Asia area. In Aikhenvald, Alexandra & R.M.W. Dixon (eds.), *The Cambridge handbook of linguistic typology*, 601-623. Cambridge: Cambridge University Press.
- Ernestus, Mirjam & R. Harald Baayen. 2011. Corpora and exemplars in phonology. In Goldsmith, J. A., J. Riggle, & A. C. Yu (eds.), *The handbook of phonological theory* (2nd ed.), 374-400. Oxford: Wiley-Blackwell.
- Gahl, Susanne & Alan C. L. Yu (eds.) 2006. Special theme issue: Exemplar-based models in linguistics. *The Linguistic Review* 23, 213-379.
- Gerner, Matthias. 2009. Deictic Features of Demonstratives: A Typological Survey with Special Reference to the Miao Group. *Canadian Journal of Linguistics* 54(1), 43-89
- Givón, Talmy. 1991. The evolution of dependent clause morpho-syntax in Biblical Hebrew. In Traugott, Elizabeth C. & Bernd Heine (eds.), *Approaches to grammaticalization, types of grammatical markers, vol. 2*, 257-310. Amsterdam: Benjamins.
- GMYZWY = Guizhousheng Minzu Yuwen Zhidao Weiyuanhui Yanjiushi [Research Unit of the Guizhou Provincial Ethnic Languages Guidance Committee] (ed.). 1965. *Miaohan jianming cidian*. [Miao-Chinese concise dictionary.] Guiyang: Guizhou Nationalities Press.

- GMYZWY & ZKYSMYDDEGD = Guizhousheng Minzu Yuwen Zhidao Weiyuanhui Yanjiushi Chuanqiandian Fangyan Zu [Research Unit of the Guizhou Provincial Ethnic Languages Guidance Committee, Chuanqiandian Dialect Group] & Zhongguo Kexue Yuan Shaoshu Minzu Yuyan Diaocha Di Er Gongzuo Dui [Academia Sinica Minority Nationalities Language Survey, Second Workgroup] (eds.). 1958. *Chuanqiandian Fangyan Miao-Han Jianming Cidian* [Miao-Chinese Concise Dictionary, Chuanqiandian Dialect]. Guiyang: Guizhou Nationalities Press.
- Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Himmelman, Nikolaus P. 2004. Lexicalization and grammaticalization: opposite or orthogonal? In Bisang, Walter, Nikolaus P. Himmelman & Björn Wiemer (eds.), *What makes grammaticalization? – a look from its fringes and its components*, 21-42. Berlin: Mouton de Gruyter.
- Iwasaki, Shoichi & Preeya Ingkaphirom. 2005. *A reference grammar of Thai*. Cambridge: Cambridge University Press.
- Jarkey, Nerida. 2015. *Serial verbs in White Hmong*. Leiden: Brill.
- Johnson, Charles & Se Yang. 1992. *Dab neeg Hmoob: Myths, legends and folktales from the Hmong of Laos*. St. Paul, Minnesota, USA: Macalester College Linguistics Department.
- Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Johnson, Keith & John W. Mullennix (eds.), *Talker Variability in Speech Processing*, 145-165. San Diego: Academic Press.

- Johnson, Keith. 2007. Decisions and Mechanisms in Exemplar-based Phonology. In Sole, Maria-Josep, Patrice Speedor Beddor & Manjari Ohala (eds.), *Experimental Approaches to Phonology*, 25-40. Oxford: Oxford University Press.
- Kirchner, Robert. 2012. Modeling exemplar-based phonologization. In Cohn, Abigail C., Cécile Fougeron, and Marie Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*, 332–44. Oxford: Oxford University Press.
- Kuryłowicz, Jerzy. 1965. The evolution of grammatical categories. *Diogenes*, 55–71. Reprinted 1975: *Esquisses linguistiques II*, 38-54. Munich: Fink.
- Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Li, Yunbing. 2015. *Miaoyu dongci de jufa yuyi shuxing yanjiu* [A study on the syntactic semantic properties of Hmong language]. Beijing: China Social Sciences Press.
- Lyman, Thomas Amis. 1979. *Grammar of Mong Njua (Green Miao)*. Self-published.
- Meng, Chaoji. 2001. *Yaozu Bunuyu fangyan yanjiu* [Dialect studies of the Bunu language of the Yao ethnicity]. Beijing: Publishing House of Minority Nationalities.
- Mortensen, David R. 2013. Tonally conditioned vowel raising in Shuijingping Mang. *Journal of East Asian Linguistics* 22(3), 189-216.
- Mortensen, David. 2015. *soc.culture.hmong (SCH) Usenet corpus*.
<http://www.davidmortensen.org/corpora/sch_corpus-2.zip>. Self-published.
- Mottin, Jean. 1978. *Éléments de grammaire Hmong Blanc*. Bangkok: Don Bosco Press.
- Narrog, Heiko & Bernd Heine. 2018. Introduction: Typology and grammaticalization. In Narrog, Heiko & Bernd Heine (eds.), *Grammaticalization from a typological perspective*, 1-15. Oxford: Oxford University Press.

- Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.
- Parsons, P. Kenneth & R. Keith Parsons (eds.). 2001. *Hua Miao-English Glossary*. Online: <<http://www.archives.ecs.soton.ac.uk/miao/glossary/printing/ReadMe.html>>
- Payne, Thomas E. 2017. Morphological Typology. In Aikhenvald, Alexandra Y. (ed.), *The Cambridge handbook of linguistic typology*, 78-94. Cambridge: Cambridge University Press.
- Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, Joan & Paul J. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 137-158. Amsterdam: John Benjamins Publishing Company.
- Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Gussenhoven, Carlos & Natasha Warner (eds.), *Laboratory Phonology 7*, 101-140. Berlin: Mouton de Gruyter.
- Rake, Steve (ed.). n.d. A Hua-Miao archive: Songs and stories and a glossary of phrases of the Hua Miao of South West China. Online: <<http://www.archives.ecs.soton.ac.uk/miao/>>.
- Ratliff, Martha. 1990. The influence of geographical change on grammar: The case of Hmong spatial deictics. Manuscript, ICSTLL 23.
- Ratliff, Martha. 1992. *Meaningful tone: A study of tonal morphology in compounds, form classes, and expressive phrases in White Hmong*. DeKalb, USA: Northern Illinois University Center for Southeast Asian Studies.
- Ratliff, Martha. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics.
- Savina, F.M. 1916. Dictionnaire Miao-Tseu-Français. *Bulletin de l'Ecole française d'Extrême-Orient* 16, 1-246.

- Smirnova, Elena. 2015. Constructionalization and constructional change: The role of context in the development of constructions. In Barðdal, Jóhanna, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.), *Diachronic Construction Grammar*, 81-106. Amsterdam: John Benjamins Publishing Company.
- Sposato, Adam. 2015. A grammar of Xong. University at Buffalo, State University of New York Ph.D. dissertation.
- Taguchi, Yoshihisa. n.d. Vocabulary of Lan Hmyo. Excel spreadsheet.
- Taguchi, Yoshihisa. 2008. *Luobohe Miaoyu cihuishu* [A vocabulary of Luobohe Miao]. Tokyo: Tokyo Foreign Languages University.
- Taguchi, Yoshihisa. 2012. On the irrealis marker *sz*⁵⁵ in Hmyo. Paper presented at the 45th International Conference on Sino-Tibetan Languages and Linguistics, Singapore, 26–28 October 2012.
- Taguchi, Yoshihisa. 2013. On the phylogeny of Hmongic languages. Powerpoint, SEALS 23.
- Thompson, Laurence C. 1987. *A Vietnamese reference grammar*. Honolulu: University of Hawaii Press.
- Traugott, Elizabeth C. 2010. (Inter)subjectivity and (inter)subjectification: a reassessment. In Cuyckens, Hubert, Kristin Davidse & Lieven Vandelanotte (eds.), *Subjectification, intersubjectification and grammaticalization*, 29-69. Berlin: Mouton de Gruyter.
- Wang, Deguang. 1986. *Weining Miaoyu huayu cailiao* [Spoken language materials of the Weining Miao language]. *Minzu Yuwen* 1986(3), 69-80.
- Wang, Fu-Shi. 1972. The classifier in the Wei Ning dialect of the Miao language in Kweichow. *Miao and Yao linguistic studies: Selected articles in Chinese, translated by Chang Yü-*

- Hung and Cho Kwo-ray*. ed. H. Purnell, 111-185. Ithaca, NY: Southeast Asia Program, Cornell University.
- Wang, Fushi. 1985. *Miáoyǔ jiǎnzhì* [*Brief guide to Miao language*]. Beijing: Nationalities Press.
- Wedel, Andrew B. 2007. Feedback and regularity in the lexicon. *Phonology* 24, 147-185.
- White, Nathan M. n.d. Chuanqiandian corpus. Unpublished text corpus.
- White, Nathan M. 2014. Non-spatial setting in White Hmong. Trinity Western University Master's thesis.
- White, Nathan M. 2020. Word in Hmong. In Alexandra Y. Aikhenvald, R. M. W. Dixon & Nathan M. White (eds.), *The phonological word and grammatical word: A cross-linguistic typology*. Oxford: Oxford University Press.
- White, Nathan M. 2021. Prehistory of verbal markers in Hmong: What can we say? In Luca Ciucci (ed.), *From fieldwork to reconstruction: historical issues in hotspots of linguistic diversity*. A special issue of *Studia Linguistica*.
- Wu, Zhengbiao & Guangying Yang. 2010. Jian tan Miao zu yingxiong shishi “Ya Lu Wang” de ji yi zhengli wenti. [Simultaneously discussing the transcription of the Miao people's epic poem “King Ya Lu” and sorting out issues.] *Minzu Fanyì* 2010(3): 58-65.
- Xiong, Jay. 2017. *Lub Hmoob txhais* (*Hmong-English dictionary*). <<http://www.hmongdictionary.com>>.
- Xiong, Yuyou & Diana Cohen. 2005. *Student's practical Miao-Chinese-English handbook*. Kunming, China: Yunnan Nationalities Publishing House.
- Yip, Po-Ching & Don Rimmington. 2004. *Chinese: A comprehensive grammar*. Abingdon, UK: Routledge.

Abbreviations

1 first person

2	second person
3	third person
A	transitive subject
ANIM	animate
CC	copula complement
CFACT	contingent fact
CLF	classifier
CoCl	complement clause
COMP	complementizer
COP	copula
CQD	Chuanqiandian
CS	copula subject
DEM	demonstrative
E	extended argument
FWH	Far Western Hmongic
GENER	general
INDF	indefinite
INTS	intensification
IRR	irrealis
LOCL	localizer
MED	medial
MH	Mile Hmong
NEG	negation
O	transitive object
PA	non-singular/abstract
PH	peripheral argument
PL	plural
RC	relative clause
REDUP	reduplicant
REL	relative clause marker
S	intransitive subject
SCH	soc.culture.hmong
SMC	Standard Mandarin Chinese
SG	singular
TOP	topic
WA	Weining Ahmao
WH	White Hmong

Figures

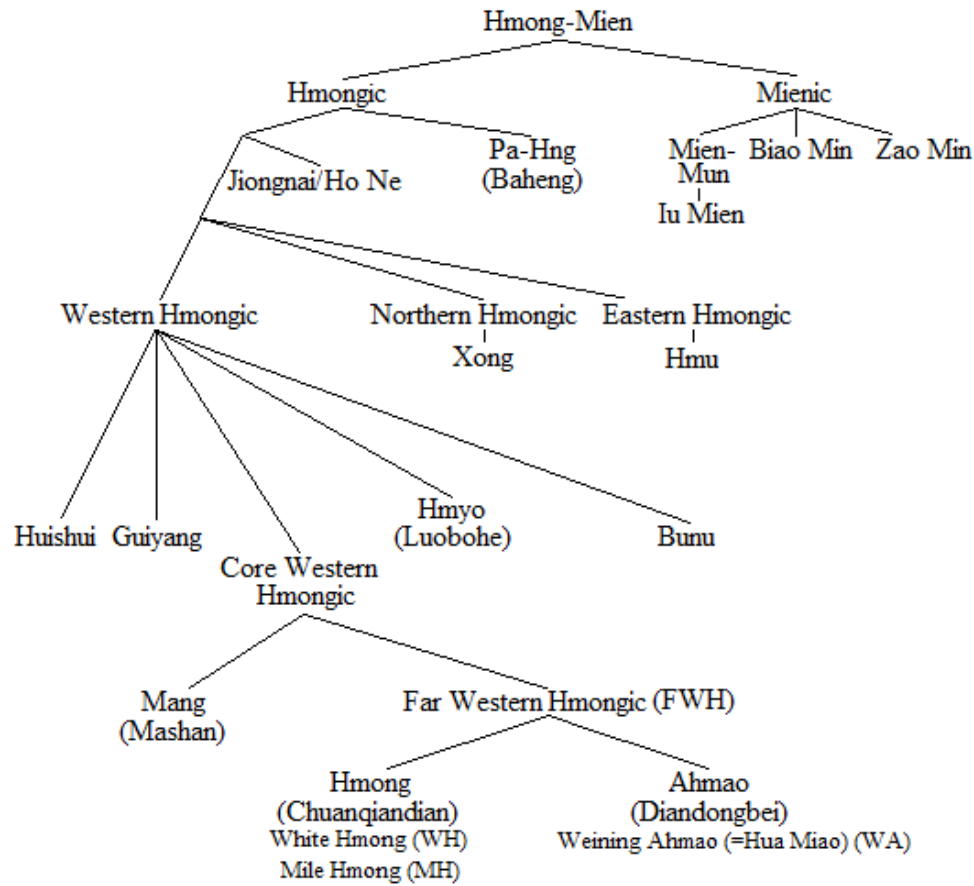


Figure 1. Hmong-Mien family tree (adapted from Ratliff 2010, Mortensen 2013 and Meng 2001).¹¹

¹¹ Note that this family tree is simplified to a significant extent, but exhibits the relationships necessary for the discussion below.

Tables

	labial	dental/alveolar	palato-alveolar	retroflex	palatal	velar	uvular	glottal
stop	/p/ <p>	/t/ <t> /?d/ <d>		/ʈ/ <r>	/c/ <c>	/k/ <k>	/q/ <q>	/ʔ/ <>
prenasalized stop	^m p/ <np>	ⁿ t/ <nt>		ⁿ ʈ/ <nr>	ⁿ c/ <nc>	ⁿ k/ <nk>	ⁿ q/ <nq>	
aspirated stop	/p ^h / <ph>	/t ^h / <th> /d ^h / <dh>		/ʈ ^h / <rh>	/c ^h / <ch>	/k ^h / <kh>	/q ^h / <qh>	
prenasalized aspirated stop	^m p ^h / <nph>	ⁿ t ^h / <nth>		ⁿ ʈ ^h / <nrh>	ⁿ c ^h / <nch>	ⁿ k ^h / <nkh>	ⁿ q ^h / <nqh>	
stop with lateral release	/p ^l / <pl>							
prenasalized stop with lateral release	^m p ^l / <npl>							
aspirated stop with lateral release	/p ^{lh} / <plh>							
prenasalized aspirated stop with lateral release	^m p ^{lh} / <nplh>							
voiceless fricative	/f/ <f>	/s̺/ <x>	/ʃ/ <s>		/ç/ <xy>			/h/ <h>
voiced fricative	/v/ <v>		/ʒ/ <z>		/j/ <y>			
affricate		/tʃ/ <tx>	/tʃ̺/ <ts>					
prenasalized affricate		ⁿ tʃ/ <ntx>	ⁿ tʃ̺/ <nts>					
aspirated affricate		/tʃ ^h / <txh>	/tʃ̺ ^h / <tsh>					
prenasalized aspirated affricate		ⁿ tʃ ^h / <ntxh>	ⁿ tʃ̺ ^h / <ntsh>					
voiceless nasal	/m̥/ <hm>	/p̥/ <hn>			/ɲ/ <hny>			
voiceless nasal with lateral release	/m̥ ^l / <hml>							
voiced nasal	/m/ <m>	/p/ <n>			/ɲ/ <ny>	(/ŋ/ <g>)		
voiced nasal with lateral release	/m ^l / <ml>							
voiceless liquid		/l̥/ <hl>						
voiced liquid		/l/ <l>						

Table 1. Consonants in White Hmong (with one marginal phoneme in parentheses).

	front	non-front
high	/i/ <i>	/i/ <w>
mid	/e/ <e>	/e/ <u>
mid nasalized	/ẽ/ <ee>	/ẽ/ <oo>
low	/a/ <a>	/v/ <o>

Table 2. Vowel phonemes in White Hmong.

Diphthong	Description
/aĩ/ <ai>	front rising
/œ/ <au>	back rising
/vi/ <aw>	mid rising
/œ̥/ <ua>	back-to-mid offglide
/i̥/ <ia>	front-to-mid offglide

Table 3. Diphthong phonemes in White Hmong.

Orthographic representation	Description	Numeric representation
	high level	55
<j>	high falling	53
<v>	mid rising	24
<s>	below-mid level	22
<> ¹²	mid level	33
<g>	falling breathy	42 (male)/53 (female)
<m>	low falling creaky	31

Table 4. Phonemic tones in Hmong.

Author Contact Information

Nathan M. White

Postal address:

Language and Culture Research Centre

Building D3, James Cook University

14-88 McGregor Rd

Smithfield QLD 4878

Australia

E-mail address: nathan.white1@my.jcu.edu.au

¹² Empty brackets indicate that the element is represented in the orthography by the absence of a symbol.