

# Re-assessing tonal diversity and geographical convergence in Mainland Southeast Asia

Marc Brunelle, University of Ottawa  
James Kirby, University of Edinburgh

## 1. Introduction: Tone typology and contact-induced tonogenesis

Mainland Southeast Asia (MSEA) is often described as the quintessential *Sprachbund*, or language area, in which languages belonging to different language families converge as a result of contact (Alieva 1984; Enfield 2005). While we hold this to be true in a general sense, we suspect that there is little to be gained in arguing about what defines a language area or in determining the exact boundary of this language area (e.g., should it just include the mainland or insular Southeast Asia as well?). What seems much more interesting to us is to gain a better understanding of how convergence happens for specific features, especially phonological and phonetic features. In this paper, we look in detail at a specific phonological feature, tone, and at two of its phonetic correlates, pitch and voice quality. Based on a database of 197 languages and dialects (§2), we assess the extent of tonal diversity in MSEA languages (§3) and construct a statistical model of the degree to which tonal inventories can be predicted on the basis of geographic proximity, genealogical relatedness and population size (§4).

Although it is generally agreed that MSEA languages are highly tonal, this characterization is often based on large national languages. Furthermore, there is often little attention paid to the types of phonetic properties that characterize tonal inventories. To our knowledge, the only systematic attempt to establish a topography of tone in Southeast Asia is Henderson (1965), who looked, among other features, at lexically contrastive pitch, phonation type, and combinations thereof. In this study, Henderson showed convincingly that tone is more prevalent on the mainland than in the archipelago and that phonation type plays a crucial role in MSEA lexical contrasts. However, because of the state of the field in 1965, Henderson's observations were only based on 31 MSEA languages, and she had limited access to phonetic data.

One motivation for the current study is to reassess Henderson's results based on an expanded sample of languages. Just like her, we have decided to focus on two phonetic properties of lexical tone: pitch, the usual suspect, and phonation type, which we will call voice quality. Other properties, such as rhyme duration, intensity and vowel quality, should ultimately be considered as well, but had to be left out as most existing phonetic/phonological studies of MSEA languages do not describe their tonal systems to this level of phonetic detail.

The second issue we address in this paper is the role of contact in tonogenesis. To our knowledge, the first piece of scholarly work that explicitly and comprehensively tackles the issue is Matisoff (1973). On the one hand, Matisoff recognizes that some languages are tone-prone, i.e. have structural characteristics that favor tonal development, like the loss of laryngeal contrasts

(well established since Haudricourt 1954) or a trend towards monosyllabization. However, he also considers contact an important driving force. More specifically, Matisoff sees Chinese influence as a crucial factor in the development of tones in Tai-Kadai, Miao-Yao and Vietnamese:

“It seems likely that the development of true tones in Vietnamese was precipitated not only by influence from Chinese, but also from Siamese as well. This indicates that Tai (and Miao-Yao) acquired their tone systems from Chinese before Vietnamese did...” (Matisoff 1973: 88).

This scenario is not without problems, however. How does a language “acquire” tone from a neighbor, even under intense contact? Pulleyblank (1986), who generally agrees with Matisoff, states the problem in the following way:

“How such a trend [i.e. tonogenesis] can spread across linguistic boundaries is an intriguing puzzle, on which I shall not venture to make any guesses” (Pulleyblank 1986: 78).

To our knowledge, the only attempt to tackle this issue in more detail is Ratliff (2002), who proposes that Proto-Hmong-Mien must have borrowed Chinese loanwords at a time when neither languages were tonal, and that they must have undergone tonogenesis in parallel. However, Ratliff generally accepts the view that contact played a role in the development of tonality in the Sinosphere.

More recently, there has been a proliferation of case studies suggesting that languages belonging to originally atonal families have or are developing tones under the influence of Tai-Kadai and Vietnamese. The following quote, chosen for its careful wording, illustrates the type of processes that might be at play:

“Experimental findings and impressionistic observations imply that both languages, Suai and Pattani Malay, are pursuing different paths leading to phonological shifts from clear and breathy voice registers for the former, and the latter, from word-initial distinctive consonant length, to a kind of prosodic salience. This could be a matter of a replacement by phonemic stress or accent, yet, given the close contact with Thai, a tone language, and the widespread bilingualism of the speakers of the two minority languages, we may have here a way station on the road to tonogenesis.” (Abramson 2004)

The large-scale scenario that emerges from these accounts is that tone would have first developed in Chinese two thousand years ago. Tone-prone languages spoken in the southern Sinosphere, like the ancestors of Tai-Kadai, Miao-Yao and Vietnamese, would then have acquired tone under Chinese influence; contact with their modern daughters, such as Thai, Lao and Vietnamese, would similarly explain tonogenesis in smaller languages. This scenario suggests that tone convergence due to language contact is a force that has been affecting the shape of tone systems for centuries.

The role of this constant force in the distribution of tone languages in modern-day MSEA will be tested in §4.

## 2. The database

In order to support a quantitative analysis of tonal convergence, we constructed a database of 197 MSEA languages. Languages were included in the database if reliable descriptions were available, if they were spoken in one of the eight MSEA countries (Vietnam, Laos, Cambodia, Thailand, Burma, Malaysia and Singapore) and if they belonged to one of the five MSEA language families (Austroasiatic, Austronesian, Sino-Tibetan, Tai-Kadai and Hmong-Mien). For languages spoken in several countries, one variety per country was included in the database, as long as data was available about its tone system (for example, Mon is counted twice, as a language of Burma and as a language of Thailand). If a language has several varieties with different tone systems in the same country, all varieties for which data was available were included (for example, northern and southern Vietnamese are counted as two distinct varieties, along with other regional dialects, as they have different tone systems). Linguistic and geographical data were extracted from available descriptions. Population figures were based on national census figures, on *Ethnologue* or, when available, on information contained in linguistic descriptions. More details and proper acknowledgements are given in the database (available upon request).

For each variety included in the database, the following information was included:

### Geographical, demographic and genealogical factors

- Language family: 87 Austroasiatic varieties (all Mon-Khmer), 19 Austronesian varieties, 40 Sino-Tibetan varieties, 43 Tai-Kadai varieties and 8 Hmong-Mien varieties. We have not subdivided languages into smaller groupings because of disagreement about subgroupings in some families and because our language sample is too small to create statistically meaningful subgroups.
- Population of the described variety, in number of speakers.
- The specific location, in longitude and latitude, of the variety described in the scholarly materials used in building the database. Whenever descriptions of varieties spoken at different locations were available, we chose to report the largest community.
- Total population of all the varieties of the same language, in number of speakers.

### Phonological variables

- Number of contrastive tones: the number of lexically contrastive categories distinguished by differences in pitch and/or voice quality on the syllable bearing the largest number of such contrasts. As an illustration, Northern Vietnamese, which has six contrastive tones in open syllables and two contrastive tones on checked syllables, can be analyzed as having six or eight tones. Here, we assume that the two checked tones can be analyzed as allotones of two of the open tones and settle for a six tone analysis.

- Number of pitch units: the number of different pitch curves used to distinguish the contrastive tones described above, even if they are redundant with voice quality (as few descriptions provide this type of information or state explicitly which cue is primary).
- Number of voice qualities: the number of different voice qualities used to distinguish the contrastive tones described above, even if they are redundant with pitch (once again, few descriptions provide this type of information or state explicitly which cue is primary).
- Word type: The maximal “non-marginal” phonological stem after excluding Western, Pali and Sanskrit loanwords. There are three possible categories: monosyllabic, sesquisyllabic and polysyllabic.

To avoid confusion, we use the term *tone* only when referring to lexically contrastive units, whereas pitch units and voice quality are often redundant and thus cannot be characterized as contrastive or not. For instance, Mon-Khmer register languages typically have a contrast between a modal/high-pitched and a breathy/low-pitched register. In most cases, we do not know what is the primary contrastive cue and what is the redundant one, notwithstanding the fact that these two cues may not be fully distinct perceptually (Brunelle 2012). A register language of that type would thus be analyzed as having two contrastive tones, two pitch units and two voice qualities.

The most difficult type of classification decision we had to make occurred when a tone system combined both contrastive pitch curves and voice quality units. We illustrate this with Northern Vietnamese, the language for which alternative classifications would yield the most important discrepancy. The Northern Vietnamese tone system has six tones in open syllables. These six tones consist of six distinct pitch curves and at least three surface voice qualities (or even more if we adopt a fine-grained typology like Nguyễn and Edmondson 1997). However, perceptual investigation reveals that not all of these properties are contrastive and that listeners seem to rely on a matrix of three relevant pitch shapes and two relevant voice qualities to distinguish the six tones (Brunelle 2009). Depending on how we count, we could therefore have five (3+2) or six tones (3×2) in Northern Vietnamese, but more importantly, we could reduce its number of pitch and voice quality units to three and two, respectively, rather than the surface six pitch contours and three voice qualities that are commonly reported in the literature. We settled on the latter option for two reasons. First, the primary materials we relied on rarely provide the level of instrumental and experimental description that would be needed to do a strictly contrastive classification. Second, most of the languages that have two possible classifications happen to be Vietic and Northern Mon-Khmer (along with a handful of Tibeto-Burman languages). Thus, the more superficial type of classification increases the tonality of many Austroasiatic languages in contact with tonal languages of other families and maximizes the probability that our models will detect a geographical convergence effect (which, as we will see in §4, we still failed to uncover.)

Classification decisions aside, some factual errors and misinterpretations of previous work are bound to have crept in. Moreover, it is possible that some of the descriptions we relied on are erroneous or tacitly avoid discussing some aspects of tone systems. Voice quality, for instance, seems to have been generally ignored in descriptions of Tai-Kadai languages until very recently. We welcome help from language specialists interested in revising parts of our database.

### 3. The typology of tone in Mainland Southeast Asia

A first look at the database, as summarized in Figure 1, reveals that contrastive tone is found in the majority of MSEA languages, but that close to 20% of the languages of the area are atonal. Another 20% have an equal number of tones and voice qualities, and could therefore be treated as *register* languages (i.e. languages in which pitch and voice quality are redundant). Therefore, depending on how we categorize languages, up to 40% of the languages of the area do not have contrastive pitch. Another interesting observation is that among languages that have contrastive tone, 66% also employ more than one type of voice quality, a proportion that reaches 54% even if we exclude register languages. Note, however, that in most of these languages, only two voice qualities, modal voicing and glottalisation/creakiness (or more rarely, breathiness), accompany the pitch-based contrast. Overall, the more pitch units a language employs, the more likely they are to be accompanied by differences in voice quality. One last observation is the relative rarity of languages with three tones, which merely reflects the history of the five language families spoken in the area: while Tai-Kadai and Hmong-Mien underwent a three-way tone split followed by a further two-way split, most of the Austronesian and Austroasiatic languages that have tonal contrasts only underwent a two-way split (Sino-Tibetan is more diverse and less reliably reconstructed). Based on the apparent cut-off at three tones, we could say that a little less than half of our sample is composed of languages that are atonal or weakly tonal, while a comparable number of languages have complex tone systems (4 tones or more).

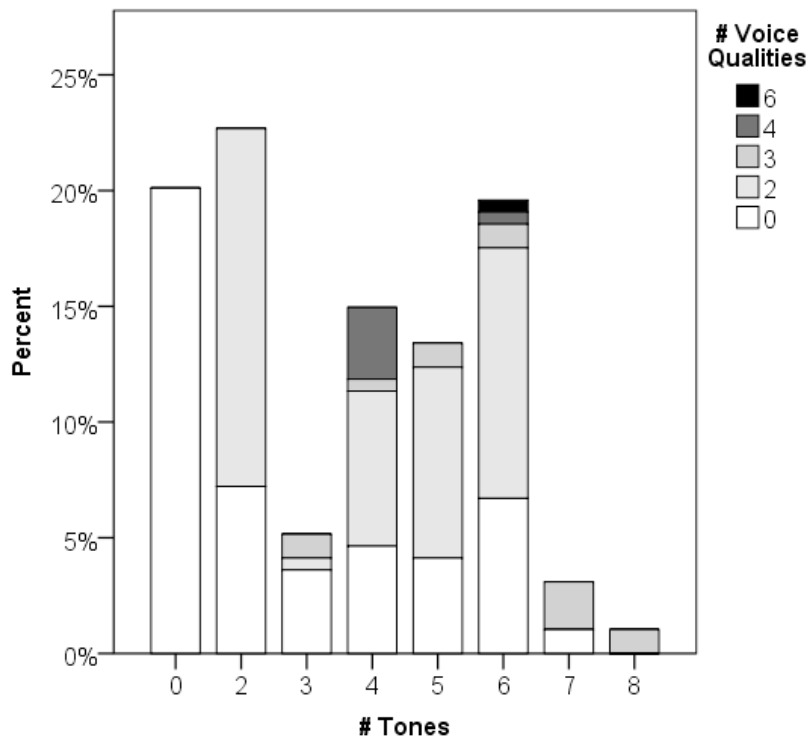


Figure 1: Co-occurrence of tone and voice quality in Mainland Southeast Asian languages

We will now look at the geographical distribution of tone, pitch and voice quality, but before doing so, a look at the geographical distribution of language families is necessary. We see in Figure 2 that only Austroasiatic and Tai-Kadai are fairly well (though far from perfectly) distributed throughout the area. Sino-Tibetan is mostly found in the northwest, Austronesian in the south, and Hmong-Mien is concentrated in the north-central zone.

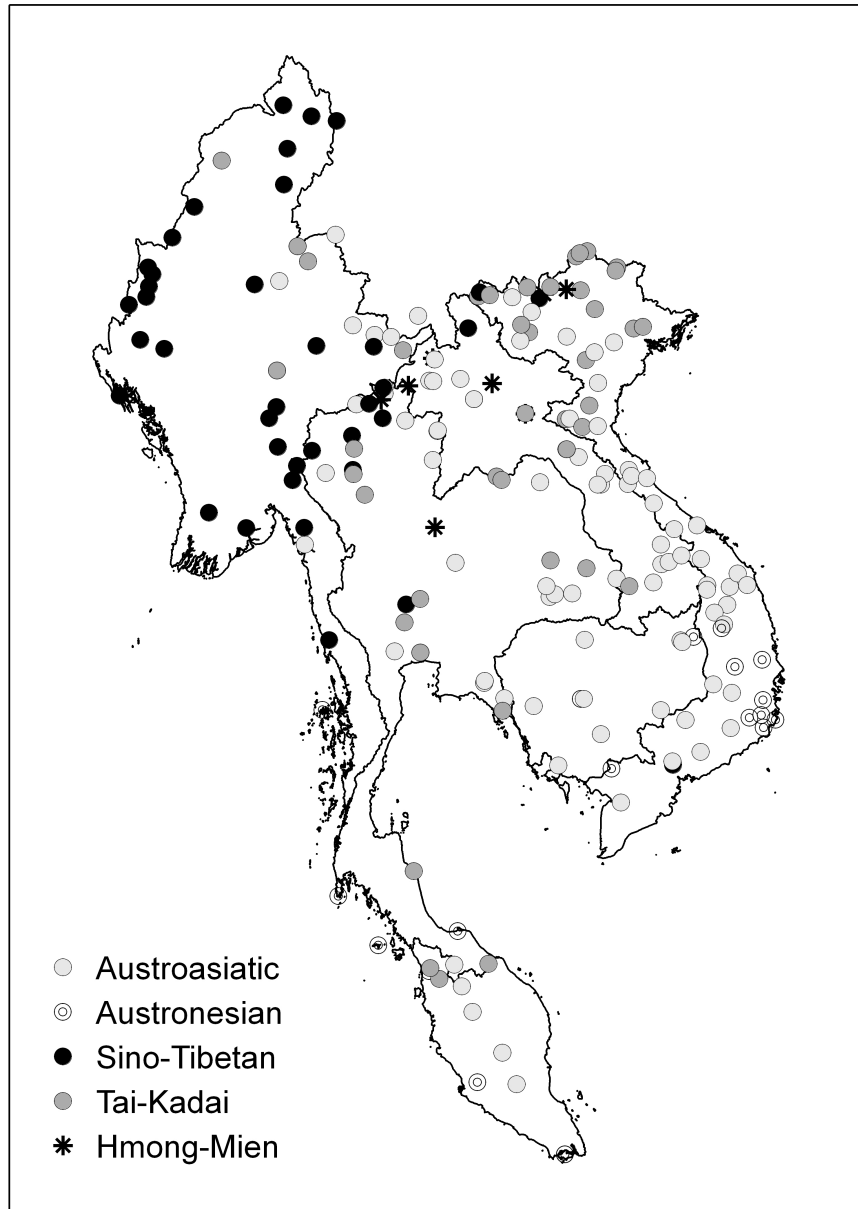


Figure 2: Geographical distribution of language families in Mainland Southeast Asia

The geographical distribution of contrastive tone in MSEA is given in Figure 3. Atonal languages are more common in the south (the Malay Peninsula and southern Vietnam), while languages with large numbers of contrastive tones tend to be found in the north. There are notable exceptions, however, like highly tonal southern Thai dialects and a few atonal Mon-Khmer languages in northern MSEA. Overall, a comparison of Figures 2 and 3 reveals that there

is a strong correlation between geography and language family: most atonal languages are Austroasiatic and Austronesian, two families that are mostly spoken in the south of the area of interest. We come back to this issue at the end of this section.

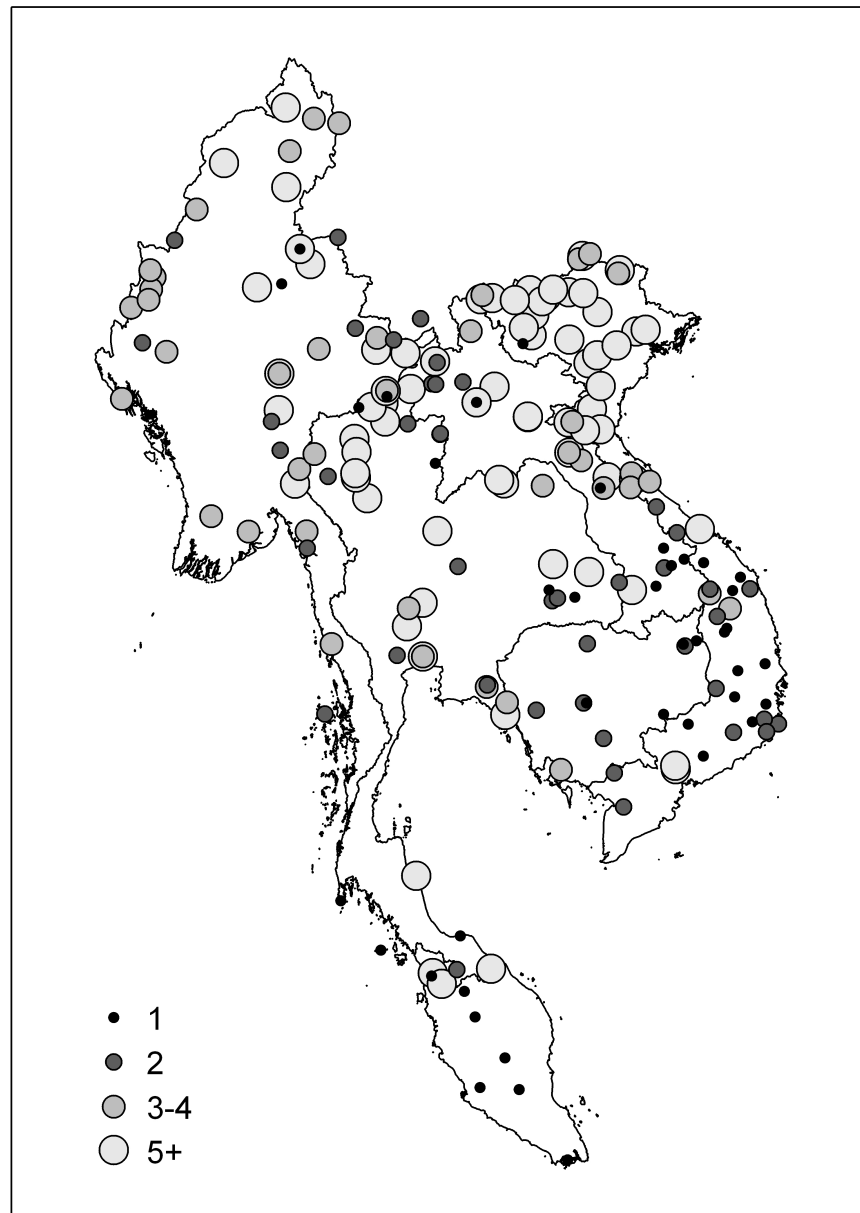


Figure 3: Number of tones per language

Figure 4 gives the geographical distribution of languages by number of pitch units. There is little difference between this map and the preceding, which simply confirms that contrastive lexical tones normally have a pitch component.

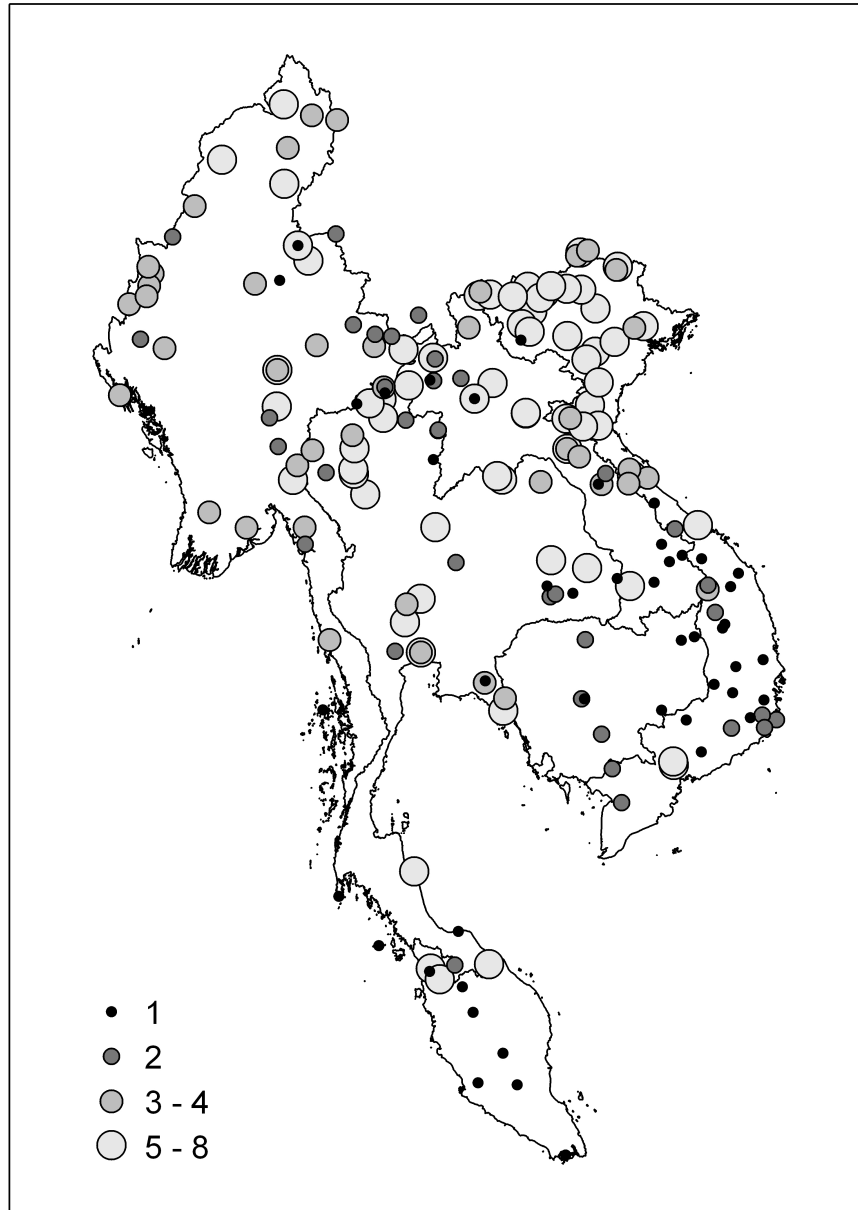


Figure 4: Number of different pitch units per language

Figure 5 finally shows the distribution of voice qualities. Languages that lack linguistically relevant voice quality are once again concentrated in southern Vietnam and the Malay Peninsula. Languages with two voice qualities are found throughout MSEA. Larger numbers of voice quality types (3-4) seem more common in the north-east, with a maximum of 6 in Kri (Enfield and Diffloth 2009). Interestingly, there does not seem to be an obvious correlation between the prevalence of voice quality and language family, something we will test in more detail in §4.



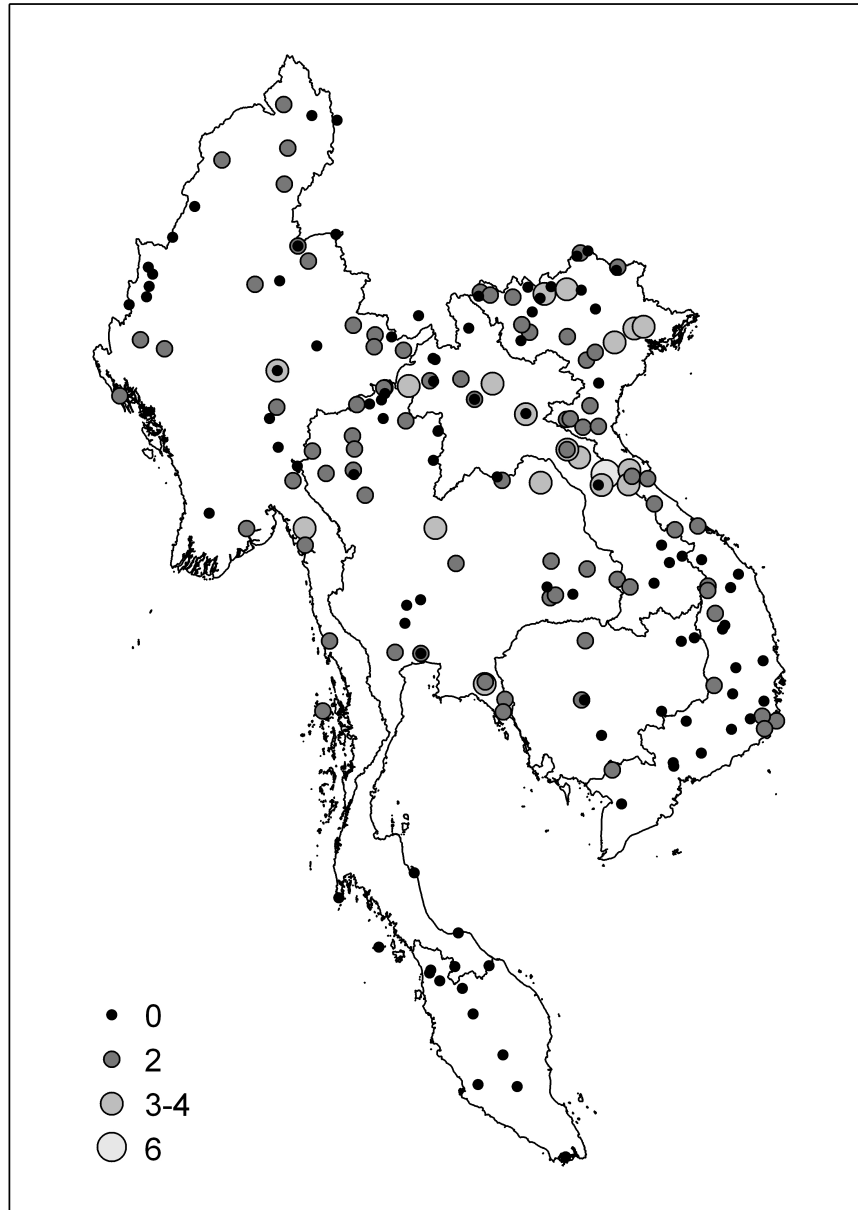


Figure 5: Number of voice qualities per language

Since geographical distribution and language family are not independent (language families are not equally distributed in MSEA), we need to look at the types of tone systems attested in the different families. Figure 6 groups tone systems into four types and gives their relative proportion in each family. We can first see that atonal languages are exclusively found in Austroasiatic and Austronesian languages. All Hmong-Mien, Tai-Kadai and Sino-Tibetan languages found in MSEA exhibit some form of tone. These results might seem trivial, but they clearly illustrate some regularities: first, the proportion of Austroasiatic languages that have developed tones is not negligible: even if we exclude register systems, which are most certainly a development internal to Mon-Khmer, 36% of Austroasiatic languages are now tonal. Second, languages that belong to families with atonal ancestors (Austroasiatic and Austronesian) can

become tonal, but languages with tonal ancestors do not lose their tones altogether. Note, however, that this directional bias is not a universal in the strong sense: register, is occasionally “restructured” into complex vowel systems in Austroasiatic and Austronesian (Huffman 1976; Lee 1977), and although we do not have cases of complete neutralization of pitch contrasts in MSEA, reductions of pitch inventories through mergers are common: for instance, Southern Vietnamese dialects have merged the tones *hỏi* and *ngã* and most Tai languages have merged some of the original six tone categories.

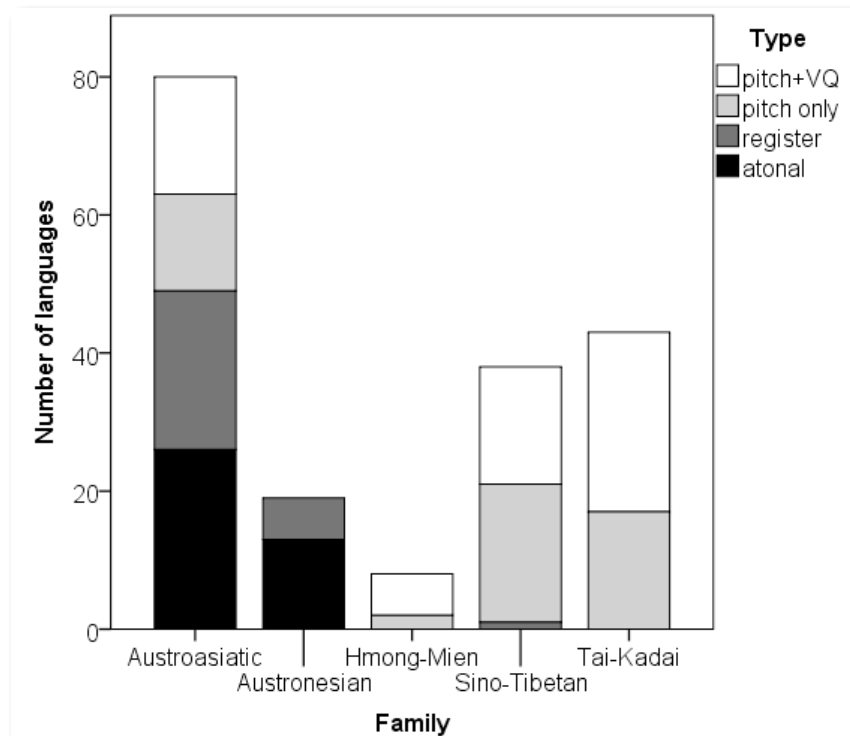


Figure 6: Tonality type, per family

#### 4. Tonality and contact-induced change in Mainland Southeast Asia

In this section, we try to determine if geographical proximity, which we use as an admittedly imperfect proxy for contact, is a predictor for the number of tones, pitch units and voice qualities found in a language. We are expecting that if contact plays a role in tonal convergence, *all other things being equal*, then neighboring languages should be more similar than distant languages. We are working on two assumptions:

- 1) The influence of the mass media and of the institutions promoting national languages (schools, military service, etc.) is recent enough that they are probably relatively ineffective in distant communities.
- 2) Population movements have been limited enough in most of the area in the recent past (since the Tai southward migrations) that conclusions based on the current geographical location of modern language communities can be projected a few centuries into the past.

We know that this assumption is inaccurate in the case of Hmong-Mien languages and of some refugee communities in Northern Thailand, but statistical tests reveal that this does not affect our results significantly.

There are also limitations to the models we are using, which are either due to the unavailability of data or to practical implementation issues:

- 1) Our measurement of geographic distance is based on specific geographic coordinates (points) rather than areas speaking the given variety (polygons). However, the geographic smooths we are using (see §4.1) are sensitive to the types of topographical features that slow down or speed up communication (e.g. mountain ranges, rivers, etc.).
- 2) Some variables, like population size, would be better factored in as relative variables defined for each pair of languages. Due to the relative sparseness of the dataset, this is not feasible using our current approach.

#### *4.1 Modeling the effects of geography with Generalized Additive Models*

How can we model differences in tonal inventories as a function of distance? Perhaps the simplest idea would be to include latitude and longitude as predictors in a linear model, but this approach places severe and wholly inappropriate restrictions on the kinds of geographic effects that can be modeled. In particular, such an approach provides no way to capture potentially local areas of tonal convergence, nor can it take into account the potentially disruptive influence of topography. If we are to take seriously the hypothesis of areal effects on tonal inventories, a more sophisticated technique will be necessary.

Here we follow recent work in statistical dialectometry (Weiling, Nerbonne et al. 2011; Wieling 2012) in making use of the generalized additive model (GAM) framework (Hastie and Tibshirani 1990; Wood 2006). A GAM is a type of statistical model very much like classical multiple linear regression, but with the ability to capture non-linearities in the way that a predictor variable influences the response variable. The use of simple linear regression with latitude and longitude as predictors would only allow us to capture hypothetical linear effects – for instance, that the degree of influence one language exerts upon another is related to the Euclidean distance between those languages, as measured by the shortest path between them on a map. While this may be true in some cases, one can easily imagine scenarios where it is inappropriate: two languages may be spoken in villages that are only a few miles apart as the crow flies, but separated by an impassable mountain range or valley. Similarly, the effects of Euclidean distance between two languages could be mitigated if they both lie along a major trade route (e.g. a river). Using GAMs, we can construct a model that is sensitive to this type of topographic variation.<sup>1</sup>

---

<sup>1</sup> While it is possible to include non-linear (parabolic or otherwise polynomial) predictors in a standard multiple regression, their shapes need to be specified in advance. The GAM framework provides a principled means of determining the shapes of these components automatically; see Wood (2006) for details.

As a first pass, we built a GAM to predict the number of tones based on the (potentially non-linear interaction of) latitude and longitude. The plot in Figure 7, created using the R package *mgcv* (Wood, 2006), shows the results. Lighter colors represents areas where the model predicts languages to have fewer tones, while darker areas represent regions of greater tonality.

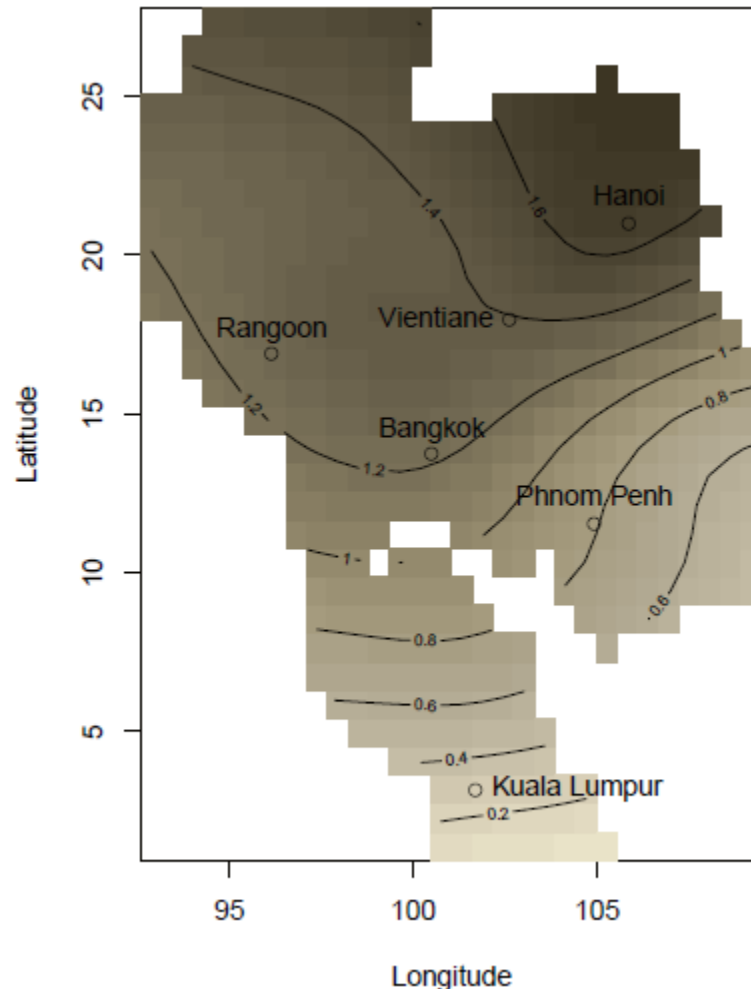


Figure 7: Contour plot of geographic effects on number of tones. Lighter grays indicate areas of fewer tones, darker grays areas with more tones. Black lines show isoglosses; the numbers indicate the logarithm of the predicted number of tones for the bounded area.

Figure 7 captures the same information as Figure 3: it indicates that, broadly speaking, the area of greatest tonality is northern Vietnam, while in the southern regions (southern Vietnam and Malay Peninsula) there are fewer tones. Importantly, there do not appear to be any dark regions surrounded by lighter ones, or vice versa, suggesting that languages in close geographic proximity tend to have similar numbers of tones (although they may still differ in other aspects of their tonal inventory). Although this might suggest a potentially strong effect of contact on tonal inventories, the following sections will demonstrate that there does not appear to be a geographical influence on the distribution of tone that is independent of language family.

## 4.2 Modeling the size of the tonal inventory

The first set of models we discuss attempt to predict the size of tonal inventories. Due to partially missing information on several languages, only 175 of the languages in the database were used as data points in the statistical analyses described below.<sup>2</sup>

### 4.2.1 Model predictors

In addition to the non-linear ‘smooth’ term representing geographic proximity (henceforth GEOGRAPHY), we considered a subset of the variables from the database described in Section 2. These included language FAMILY, the local population size (POPLOCAL) and the total population size (POPTOTAL) along with the language’s canonical WORDTYPE (mono-, sesqui-, or polysyllabic). Of these predictors, FAMILY and WORDTYPE were included as fixed-effect predictors; we also considered models where the effects of geography were potentially affected by the local and total population sizes (for instance, where a small language community has little effect on a large one, even if the two languages are spoken in the same location, or where a large population affects a small one despite of a large distance).

### 4.2.2 Results

Because our dependent variable (number of tones) takes on successive non-negative integer values (i.e. counts from 0 to 8), we assumed it to follow a Poisson distribution.<sup>3</sup> We considered a range of models, starting with a simple model containing just a single predictor (FAMILY or WORDTYPE), then adding predictors and checking if their inclusion resulted in a justified increase in model complexity.

Table 1 lists the coefficients and associated statistics of our final model, which has an adjusted  $R^2$  of 0.678. This model contains two predictors, FAMILY and WORDTYPE; the base levels are *Austroasiatic* for FAMILY and *monosyllabic* for WORDTYPE. As we used a logistic link function, the estimates are logarithms, but it is simple to transform them into integer estimates. For example, the predicted number of tones for a polysyllabic Sino-Tibetan language is  $\exp(1.3408) + \exp(0.4818) - \exp(0.4446) = 3.88$ , while for a monosyllabic Austronesian language the estimate is  $\exp(1.3408) - \exp(0.6091) = 1.98$ .

	Estimate	Std. error	z-value	p-value
(Intercept)	1.3408	0.1315	10.193	<0.0001
FAMILY=Austronesian	-0.6091	0.2173	-2.803	<0.01
FAMILY=Hmong-Mien	0.6052	0.1875	3.227	<0.01
FAMILY=Sino-Tibetan	0.4818	0.1410	3.418	<0.001

<sup>2</sup> It would have been possible to include up to 186 languages for some of the models (like those that do not factor in population), but since this has little effect on the overall results, we have favored a uniform approach that allows easy statistical comparison of the models.

<sup>3</sup> In particular, we employed generalized additive Poisson regression models using a logarithmic link function, with smoothing parameters estimated using the method of restricted maximum likelihood (REML).

FAMILY=Tai-Kadai	0.3891	0.1481	2.628	<0.01
WORDTYPE=sesquisyllabic	-0.5554	0.1414	-3.928	<0.0001
WORDTYPE=polysyllabic	-0.4446	0.1608	-2.764	<0.01

Table 1: Significant parametric coefficients and associated statistics for the final model, predicting number of tones from FAMILY (base level: Austroasiatic) and WORDTYPE (base level: monosyllabic). Estimates give the logarithm of adjustment to the intercept predicting the number of tones, with positive estimates indicating increases relative to the intercept and negative estimates indicating decreases (see text).

Despite the fact that word shapes are not evenly distributed across the five language families under consideration (see Table 2), models containing a predictor for WORDTYPE always resulted in a significant reduction in deviance. This model is consistent with the descriptive generalizations to be gleaned from our database and from previous scholarship: the number of tones tends to be inversely correlated with complexity of canonical word shape (i.e., monosyllabic languages tend to have more tones than sesquisyllabic or polysyllabic languages). As such, Hmong-Mien and Tai-Kadai languages tend to have large tonal inventories, and Austroasiatic and Austronesian languages tend to have small tone inventories (typically register systems) or to be non-tonal.

FAMILY	WORDTYPE		
	<i>monosyllabic</i>	<i>sesquisyllabic</i>	<i>polysyllabic</i>
Austroasiatic	10	62	4
Austronesian	1	10	9
Hmong-Mien	8	0	0
Sino-Tibetan	5	7	24
Tai-Kadai	34	1	0

Table 2: Distribution of word type by family.

That our best-fitting model is reasonably reflective of empirical realities allows us to have some measure of confidence in its predictions, as well as its status vis-à-vis alternative models. In particular, in none of the alternative models we considered did GEOGRAPHY emerge as a significant predictor; while a model containing GEOGRAPHY, FAMILY and WORDTYPE is not significantly worse than a model without the GEOGRAPHY smooth term (adjusted  $R^2=0.675$ ), the smooth term itself did not reach significance (see Table 3). We also considered a model in which the effect of geographic proximity was modulated by population size (POPLOCAL or POPTOTAL), in order to capture a potential asymmetry in degree of influence (whereby a language with a large population could exert a greater influence on a nearby language with a small population); however, neither model was superior to one containing a non-linear geographic predictor only.

However, as was seen in Figure 2, language families in our database are not evenly distributed throughout mainland Southeast Asia. As this is unlikely to be an artifact of our sample, we considered the possibility that GEOGRAPHY could function as an equally good predictor as FAMILY (i.e., that both variables encode similar information about tonal distributions). Indeed, in a model including just WORDTYPE and the smooth term for

GEOGRAPHY, the latter emerges as a significant predictor ( $\chi^2 = 24.88, p < 0.001$ ), but a likelihood-ratio test determines that this model is inferior to the model containing FAMILY and WORDTYPE, explaining less overall variance ( $R^2 = 0.603$  vs  $0.678$ ). Based on these results, we infer that any influence of geographic proximity on the number of tonal contrasts in a MSEA language, independent of genealogical affiliation, is likely to be *fairly small*.

Parametric coefficients:

	Estimate	Std. error	z-value	p-value
(Intercept)	1.3132	0.1413	9.296	<0.0001
FAMILY=Austronesian	-0.5136	0.2274	-2.258	<0.05
FAMILY>Hmong-Mien	0.5798	0.1978	2.931	<0.01
FAMILY=Sino-Tibetan	0.4471	0.1661	2.691	<0.01
FAMILY=Tai-Kadai	0.3676	0.1554	2.365	<0.05
WORDTYPE=sesquisyllabic	-0.5282	0.1455	-3.631	<0.001
WORDTYPE=polysyllabic	-0.3770	0.1754	-2.150	<0.05

Approximate significance of smooth terms:

	edf	Ref. df	$\chi^2$	p-value
GEOGRAPHY	2.004	2.088	2.903	0.247

Table 3: Coefficient estimates and associated statistics for a model containing the predictors GEOGRAPHY, FAMILY (base level: Austroasiatic) and WORDTYPE (base level: monosyllabic). Estimates give the logarithm of adjustment to the intercept predicting the number of tones, with positive estimates indicating increases relative to the intercept and negative estimates indicating decreases.

#### 4.3 Pitch and voice quality inventories

We also explored a number of similar models to see how well the number of pitch categories or voice qualities in a language could be predicted on the basis of the variables in our database. The same factors were tested as for the number of tones, except that the number of voice quality units (NBVQ) was also tested as a predictor for the number of pitch categories and the number of pitch categories (NBPITCH) was tested as a predictor for the number of voice qualities.

The best model for predicting pitch units includes FAMILY, WORDTYPE and NBVQ as significant predictors ( $R^2 = 0.722$ ). As we can see in Table 4, it is overall very similar to the model for tonal contrasts (Table 1), but the significant NBVQ term also shows that there is a slight positive correlation between voice quality and pitch. This reflects the common finding that voice quality tends to be accompanied by redundant pitch variations, especially in register systems. Once again, GEOGRAPHY was not a significant predictor in any of the nested models we compared.

Parametric coefficients:

	Estimate	Std. error	z-value	p-value
(Intercept)	1.13291	0.15913	7.119	<0.0001
FAMILY=Austronesian	-0.51268	0.22420	-2.287	<0.05

FAMILY=Hmong-Mien	0.42815	0.19674	2.144	<0.05
FAMILY=Sino-Tibetan	0.45891	0.14815	3.098	<0.01
FAMILY=Tai-Kadai	0.32919	0.14967	2.199	<0.05
WORDTYPE=sesqui	-0.72077	0.14509	-4.968	<0.0001
WORDTYPE=poly	-0.50725	0.16640	-3.048	<0.01
NBVQ	0.13194	0.05289	2.495	<0.05

Table 4: Significant parametric coefficients and associated statistics for the best-fit model predicting number of pitch units from FAMILY (base level: Austroasiatic), WORDTYPE (base level: monosyllabic) and NBVQ. Estimates give the logarithm of adjustment to the intercept predicting the number of pitch units, with positive estimates indicating increases relative to the intercept and negative estimates indicating decreases.

All the models we built for predicting the number of voice qualities had a low explanatory power ( $R^2$  of at most 0.23). The only interesting observation here is that in all these models, NBPITCH is the only significant factor. A look at coefficients reveals that an increase of one pitch unit results in a similar increase in number of voice qualities, which reflects the fact that languages with large numbers of pitch units are more likely to make use of voice quality in their tone systems than languages with smaller pitch inventories, all other things being equal. On the other hand, the factors that do well in predicting the number of tones or the number of pitch units (FAMILY, WORDTYPE) fail to predict the number of voice qualities found in a language.

#### 4.4 *The 'idea' of tone*

One could also imagine that contact spreads the ‘idea’ of tone, rather than influencing the number of tonal categories directly (an idea proposed by Benedict 1996 in a paper whose conclusions are otherwise not supported by our results). This would be the case if the chance of a language phonologizing previously predictable pitch variation was higher if its speakers were bilingual in a language that already made use of contrastive pitch (the same scenario may hold for voice quality). For instance, Eastern Cham, which is in contact with tonal Vietnamese, seems to have a register system that is more pitch-based (though not exclusively) than Western Cham, a language of Cambodia that is not in contact with languages making use of contrastive pitch (Brunelle 2009). Perhaps this is due to familiarity with contrastive pitch, and it is plausible that given enough time and a little chance, Eastern Cham could develop a two-way contrast based exclusively on pitch.

We explored this possibility by recoding all languages in our database as tonal or atonal. Here, we depart from the definition of tone used in the rest of the paper in that a language was designated as tonal if it employs at least two different pitch units that are not redundant with voice quality, a criterion meant to exclude from consideration canonical register systems. Table 5 shows that by this definition, the only family that exhibits any meaningful variation is Austroasiatic. Hmong-Mien is entirely tonal, Austronesian is entirely atonal, Tai-Kadai is tonal, except for Cao Lan, which is treated as an atonal register language, and the only atonal Sino-Tibetan language is Chin Daai.



Family	Number of atonal languages	Number of tonal languages
Austroasiatic	54	24
Austronesian	20	0
Hmong-Mien	0	8
Sino-Tibetan	1	36
Tai-Kadai	1	42

Table 5: Number of atonal and tonal languages per family  
(186 languages for which we have Word type data)

Since Austroasiatic is the only language that exhibits variation, a statistical model would have to factor in interactions, something for which we do not have enough observations. However, a further look at tonal Austroasiatic languages reveals that this group contains both languages which could have borrowed the idea of tone from neighboring languages, and languages that do not seem to have immediate tonal neighbors:

#### Languages with tonal neighbors

- 10 Vietic languages, 8 of which are varieties of Viet-Muong which probably have the same tonal ancestor and have not developed tones independently
- 4 Khmuic languages, including two closely-related Khmu dialects.
- 3 closely related Palaungic languages which probably have not developed tones independently
- Mang

#### Languages without tonal neighbors

- 2 varieties of Khmer that have developed marginal tone through the loss of medial /r/ (possibly a single tonogenetic event)
- 2 Bahnaric languages (Southern Jeh and Koho – the tonal status of the latter is unsure)
- 1 Aslian language (Kensiu)
- 1 Pearic language, that could probably be described as a complex register language (Samre)

Although our available data does not contain enough observations to support a robust statistical model, the hypothesis that contact spreads the idea of tone in a way independent of genealogical relationship is not clearly supported by our sample.

## 5. Discussion

On the basis of the typological survey and statistical analyses outlined above, four main findings are worth highlighting:

1. *MSEA is tonally diverse and dichotomous tonal classifications give an erroneous impression of homogeneity.* As shown in §3, 20% of the languages of the area are atonal and 20% have register systems. The 60% of tonal languages are not homogeneous but exhibit a wide range of diversity, from simple two-tone systems based exclusively on

pitch to complex tone systems combining large number of contrastive pitch units and voice qualities.

2. *The phylogenetic signal for tone is extremely strong.* In all models we considered except those accounting for the number of voice quality, FAMILY emerged as a significant predictor of the number of categories under investigation (number of tonal categories, number of pitch units, presence/absence of tone).
3. *Although FAMILY and WORDTYPE are closely related, the best models always included both factors,* indicating that both explain at least partially independent portions of the observed variance. This seem to confirm Matisoff (1973)'s view that there is a causal relation between monosyllabization and tonality. However, this causal relationship is still ill-understood. It could stem from two sources: either the loss of presyllables is accompanied by the transfer of some of their contrastive properties onto the main syllable (like the spirantization and voicing of medial obstruents in Việt-Mường described in Ferlus 1982), or monosyllables are for some reason intrinsically more likely to neutralize laryngeal contrasts in onsets and codas. Unfortunately, the former is rarely attested or reconstructed and no solid phonetic scenario seems to support the latter.
4. *When FAMILY is included in the model there is no independent effect of GEOGRAPHY.* While this could change in a model with more sophisticated geographic terms (e.g. elevation, travel distance, etc.), we suspect, based on the current results, that any influence of geographic proximity on the size of the tone inventory is likely to be extremely small.

These findings suggest to us that there is no wide scale force that directly pushes languages to become tonally similar to their neighbors. By and large, neighboring languages tend to have similar tone systems because they are related, not because they are in contact. Of course, this does not mean that there are *never* cases of contact-induced tonogenesis. Indeed there are cases where contact is the obvious explanation for tonality, but they are not cases of convergence proper. For example, in Mal, Tai loanwords bear a special tone that distinguishes them from native vocabulary (Chommanad 2010). Furthermore, as we are using synchronic tonal inventories as a proxy for the effects of contact, we might be overlooking past cases of tonal convergence. However, on the basis of the current survey, we submit that there is no evidence for broad, areal convergence of tonality, independent of genealogical affiliation.

Based on this result, what should we do with celebrated cases of contact-induced tonogenesis, like Vietnamese (Haudricourt 1954) or Tsat (Maddieson and Pang 1993; Thurgood 1993)? A first explanation is that these languages may have undergone an exceptionally intensive and long-lasting contact that goes beyond the diffuse and large-scale geographical effect measured here. We know that these conditions were probably met in the case of Vietnamese (Hastie and Tibshirani 1990), although we have little reliable information about Tsat history (Brunelle accepted). A second possibility is that some of these languages underwent tonogenesis independently of contact. This certainly happened to Chinese, allegedly the first East

Asian language to have undergone tonogenesis, and to a number of languages spoken in atonal environments, like two independent sub-groups of languages of New Caledonia (Rivierre 1993), the ancestor of Athapaskan languages (Kingston 2004), Seoul Korean (Silva 2006; Kirby 2010) and Balsas Nahuatl (Guion, Amith et al. 2010). Therefore, contact-independent tonogenesis should occasionally happen in Mainland Southeast Asia as well.

A question that immediately comes to mind and that was frequently asked by audiences when presenting preliminary version of this work is “if the role of contact is so limited why are so many MSEA languages tonal?”. While this is not a trivial question, we believe it can only be answered after considering a few issues:

- a) First of all, as shown in §3, MSEA tone systems are not homogeneous. The wide tonal diversity found in MSEA languages suggests that even if there was convergence, it would probably operate at a subtler level than a mere tonal/atonal dichotomy. Furthermore, while MSEA seems more tonal than average, there seem to be other language areas with comparable degrees of tonality. While, there is to our knowledge no other systematic survey of tone in a specific language area, a look at WALS (Maddieson 2011) suggests that West Africa, and perhaps Mesoamerica and Papua, are also highly tonal, both in the sense that tone languages are thick on the ground and that the tone systems themselves are highly complex. That MSEA is tonally exceptional in any meaningful way still remains to be demonstrated.
- b) Second, an underlying assumption behind this question is that even if there is no obvious convergence effect in modern languages, there must have been some convergence between the tone systems of the ancestors of Chinese, Hmong-Mien and Tai-Kadai (see §1 for explicit claims). While reconstructions of Proto-Tai and Proto-Hmong-Mien tone systems (Pittayaporn 2009; Ratliff 2010) suggest strong similarities with Chinese, with three tones in open syllables, it is important to insist that we know very little about the sources of these three tones and that current reconstructions do not allow us to date or locate their development. Therefore, while tone convergence in the Sinosphere cannot be excluded, it is at the moment a speculative scenario, and, in any case, its underpinnings are totally unknown. Moreover, while the five modern MSEA language families are uncontroversially recognized, it is possible that they are distantly related (see articles in Enfield 2011 for an overview of recent competing scenarios), and that tone, or tone-proneness, is partly traceable to common ancestors.
- c) Thirdly, it is important to recognize that tone languages are not present to an especially great degree in the two MSEA families that do not have tonal ancestors. MSEA Austronesian languages are either atonal or registral, and MSEA Austroasiatic languages exhibit the full range of tonal behavior (atonal, registral, pitch+voice quality, pitch only). Crucially, since Austroasiatic can probably be reconstructed with an onset voicing contrast on the verge of being phonologized into register, its descendants are expected to have evolved in various directions independently of contact, because a register contrast can easily be lost, preserved or reinterpreted as a primarily pitch contrast.

We should finally consider the possibility that contact has no direct effect on tone, but that word shape is affected by contact and in turn affects tonality (in fact, the link between word type

and tonality is supported by our results). Such scenarios would seem in line with some of the ideas raised in Matisoff (1973). Unfortunately, since the only language families that exhibit significant variation in word shape are Sino-Tibetan and Austroasiatic, a well-adapted model would require the inclusion of interactions, something for which we lack sufficient observations to test at present. More generally, while it is possible to imagine scenarios in which speakers of an atonal language become familiar with tone in their second language and then phonologized pre-existing pitch variations in their native language, it is not obvious why fluency in a monosyllabic L2 would prompt speakers to drop syllables in their native language. Complex scenarios involving simultaneous monosyllabization in two languages are possible (with the possible involvement of stress-shifts: Donegan and Stampe 2004; Brunelle and Pittayaporn 2012), but they do not explain why a polysyllabic/sesquisyllabic language would become monosyllabic after coming in contact with an already monosyllabic language. There are obviously interesting questions and possible answers here, but they are beyond the scope of this paper.

## **6. Conclusion**

The database presented in this study allows us to obtain a finer-grained typology than that presented in Henderson (1965). Pitch and voice quality are both important tonal cues in MSEA languages, but the tone systems that result from the combination of these cues are very diverse, ranging from register systems (simple or complex) to tone systems based on either only pitch or a combination of pitch and voice quality. Moreover, contrary to stereotypical views, a significant proportion of MSEA languages are atonal: 20% of the languages in our database are atonal, and depending on how we treat voice register languages, this figure could go up to 40%. The geographical distribution of tonality is also clearly skewed. Languages tend to have the most tones (and pitch units) in northern MSEA (especially Northern Vietnam) and least in southern Vietnam and Peninsular Malaysia, with a smooth gradient in between. Interestingly, this geographical effect is not statistically significant and seems to be an artifact of linguistic affiliation and word shape.

Voice quality on the other hand, seems largely random. Language family and word shape do not account for it, nor does geography. The only factor that explain the presence of voice quality distinctions, even if weakly, is the number of pitch units: languages that have many pitch categories, tend to accompany them with voice quality distinctions, probably for reasons of contrast maximization.

The second type of conclusion reached in this paper is essentially a series of negative results. While these may appear unexciting, they are nonetheless important as they challenge current views of contact and change. First of all, it seems that population size does not have a clear effect on the tendency that a language has to look like its neighbors. Thus it cannot be said that small languages are more prone to become tonally similar to their neighbors. Second, the absence of geographical effect suggests that there is no systematic large-scale effect of contact on tone inventories: all else being equal, unrelated neighboring languages do not tend to have tonal

inventories of similar size. Similarly, languages do not seem to borrow the idea of tone from their neighbors: that is, the likelihood of a language phonologizing pitch variations into tones does not seem to depend on the tonality of neighboring languages. Our results suggest a single clear trend in terms of tone change: in MSEA, atonal languages can become tonal, but tonal languages rarely become atonal (and the few attested cases belong to the subset of register languages). In the end, only two factors predict the degree of tonality of MSEA languages: the language family to which they belong, which suggests that tonality is largely inherited, and word shape (independently of family). While this finding suggests to us a potentially more complicated scenario in which geographic proximity exerts an indirect influence on tonality, the mechanism(s) driving word shape convergence would need to be spelled out in more detail before assessing such a proposal.

We would like to conclude by insisting that our models should not be interpreted as evidence that there has *never* been any form of contact-induced (or contact-favored) tonogenesis in MSEA. It is possible that such effects occur at a very local level (although an inspection of residuals from our statistical models did not reveal any such effect), or that only the inclusion of sociolinguistic factors like intensity or duration of contact would allow them to emerge. We believe that the burden of the proof falls on the proponents of such effects (we are planning to test such models ourselves), and think that it is worth insisting that since tonogenesis does occur in languages that are not in contact with tonal languages, it would be expected to occasionally occur in languages that have tonal neighbors *even if contact were to play no role at all*.

## 7. Acknowledgments

We would like to thank Martijn Weiling and Dan Dediu for their thoughts on our implementation of the statistical models, audiences at the MPI workshop and at BLS 39 for their feedback (special thanks to James Matisoff, whose ideas we seem to keep returning to), Pittayawat Pittayaporn for his help with the coding of the Tai-Kadai languages in the database, and Phạm Thị Thanh Hiền for her help with cartographical software and geographic projections.

## 8. References

- Abramson, A. (2004). Towards prosodic contrast: Suai and Pattani Malay. Proceedings of the International Symposium on tonal aspects of languages: Emphasis on tone Social Sciences. B. Bel and I. Marlien. Beijing, The Institute of Linguistics, Chinese Academy of Social Sciences: 1335.
- Alieva, N. F. (1984). "A Language-Union in Indo-China." Asian and African Studies XX: 11-22.
- Benedict, P. K. (1996). "Interphyla flow in Southeast Asia." The Fourth International Symposium on Language and Linguistics: 1579-1590.
- Brunelle, M. (2009). "Contact-induced change? Register in three Cham dialects." Journal of Southeast Asian Linguistics 2: 1-22.
- Brunelle, M. (2009). "Tone perception in Northern and Southern Vietnamese." Journal of Phonetics 37: 79-96.

- Brunelle, M. (2012). "Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham." Journal of the Acoustical Society of America **131**(4): 3088-3102.
- Brunelle, M. (accepted). Revisiting the expansion of the Chamic language family: Acehnese and Tsat. New Research in Cham Studies: an International Conference. A. Griffiths, A. Hardy and G. Wade. Singapore, Institute of Southeast Asian Studies.
- Brunelle, M. and P. Pittayaporn (2012). "Phonologically-constrained change: The role of the foot in monosyllabization and rhythmic shifts in Mainland Southeast Asia." Diachronica **29**(4): 411-433.
- Chommanad, I. (2010). "Tai loanwords in Mal: A minority language of Thailand." Mon-Khmer Studies **39**: 123-136.
- Donegan, P. and D. Stampe (2004). Rhythm and the Synthetic Drift of Munda. The yearbook of South Asian languages and linguistics. R. Singh. New Delhi, Thousand Oaks: 3-36.
- Enfield, N. J. (2005). "Areal Linguistics and Mainland Southeast Asia." Annual Review of Anthropology **34**: 181-206.
- Enfield, N. J. (2011). Dynamics of Human Diversity, Pacific Linguistics, School of Culture, History and Language, College of Asia and the Pacific, Australian National University.
- Enfield, N. J. and G. Diffloth (2009). "Phonology and sketch grammar of Kri, a Vietic language of Laos." Cahiers de linguistique - Asie orientale **38**(1): 3-69.
- Ferlus, M. (1982). "Spirantisation des obstruantes médiales et formation du système consonantique du vietnamien." Cahiers de linguistique - Asie orientale **11**: 83-106.
- Guion, S. G., J. D. Amith, et al. (2010). "Word-level prosody in Balsas Nahuatl: The origin, development, and acoustic correlates of tone in a stress accent language." Journal of Phonetics **38**(2): 137-166.
- Hastie, T. and R. Tibshirani (1990). Generalized Additive Models, Chapman and Hall.
- Haudricourt, A. (1954). "De l'origine des tons en viêtnamien." Journal Asiatique **242**: 69-82.
- Henderson, E. (1965). "The Topography of Certain Phonetic and Morphological Features of Southeast Asian Languages." Lingua **15**: 400-434.
- Huffman, F. (1976). "The register problem in fifteen Mon-Khmer languages." Oceanic Linguistics special publication Austroasiatic Studies, part 1(13): 575-589.
- Kingston, J. (2004). The Phonetics of Athabascan Tonogenesis. Athabascan Prosody. K. Rice and S. Hargus. Amsterdam, John Benjamins: 137-184.
- Kirby, J. (2010). Cue selection and category restructuring in sound change. Ph.D., University of Chicago.
- Lee, E. W. (1977). Devoicing, Aspiration, and Vowel Split in Haroi: Evidence for Register (Contrastive Tongue-Root Position). Papers in Southeast Asian Linguistics no.4. D. D. Thomas, E. W. Lee and N. Đ. Liêm. Canberra, Australian National University. **48**: 87-104.
- Maddieson, I. (2011). Tone. The World Atlas of Language Structures Online. M. S. Dryer and M. Haspelmath. Munich, Max Planck Digital Library. **feature 13A**.
- Maddieson, I. and K.-F. Pang (1993). Tone in Utsat. Tonality in Austronesian Languages. J. Edmondson and K. Gregerson. Honolulu, U of Hawaii Press: 75-89.
- Matisoff, J. (1973). Tonogenesis in Southeast Asia. Consonant Types and Tone. L. Hyman. Los Angeles, USC. **1**: 71-96.
- Nguyễn, V. L. and J. Edmondson (1997). "Tones and voice quality in modern northern Vietnamese: Instrumental case studies." Mon-Khmer Studies **28**: 1-18.

- Pittayaporn, P. (2009). Phonology of Proto-Tai. Ph.D., Cornell.
- Pulleyblank, E. G. (1986). "Tonogenesis as an index of areal relationships in East Asia." Linguistics of the Tibeto-Burman Area **19**(1): 65-82.
- Ratliff, M. (2002). "Timing tonogenesis: Evidence from borrowing." Proceedings of the Twenty-Eight Annual Meeting of the Berkeley Linguistics Society: Special Session on Tibeto-Burman and Southeast Asian Linguistics: 29-41.
- Ratliff, M. (2010). Hmong-Mien Language History. Canberra, Pacific Linguistics.
- Rivierre, J.-C. (1993). Tonogenesis in New Caledonia. Tonality in Austronesian Languages. J. A. Edmondson and K. Gregerson. Honolulu, University of Hawai'i Press: 155-173.
- Silva, D. J. (2006). "Acoustic evidence for the emergence of tonal contrast in contemporary Korean." Phonology **23**(02): 287-308.
- Thurgood, G. (1993). Phan Rang Cham and Utsat: Tonogenetic Themes and Variants. Tonality in Austronesian Languages. J. Edmondson and K. Gregerson. Honolulu, U of Hawaii Press: 91-106.
- Weiling, M., J. Nerbonne, et al. (2011). "Quantitative social dialectology: Explaining linguistic variation geographically and socially." PLOS ONE **6**(9 e23613).
- Wieling, M. (2012). A quantitative approach to social and geographical dialect variation, Groningen Dissertations in Linguistics 103.
- Wood, S. (2006). Generalized Additive Models: An Introduction with R, Chapman & Hall.