

# AREAL TYPOLOGY OF MAINLAND SOUTHEAST ASIA: WHAT WE LEARN FROM THE *WALS* MAPS

Bernard Comrie<sup>1</sup>

## Abstract

*Mainland Southeast Asia has long been recognized as a classic example of a linguistic area, but earlier characterizations of this language area have typically been intuitive, for instance providing seemingly impressive lists of features known to be shared by Mainland Southeast Asian languages but without considering a list of features on which these languages differ, without explicitly considering the extent to which the features in question are common or rare across the world as a whole. By using the maps in the *World Atlas of Language Structures*, it is possible to build up a more structured assessment of the extent to which Mainland Southeast Asia constitutes a linguistic area. Many maps show a clear delimitation between Mainland Southeast Asia and the rest of Eurasia, although the precise boundary varies from map to map, as does the presence and location of intermediate zones. The dividing line between Mainland Southeast Asia and Insular Southeast Asia is much less clear-cut, thus providing some evidence for a more general Southeast Asian linguistic area.*

---

<sup>1</sup> Professor and Director, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz, 6 D-04103 Leipzig Germany  
comrie@eva.mpg.de

## Introduction

Mainland Southeast Asia is often cited as one of the best, if not the best, examples of a linguistic area, i.e. as an area that is characterized on the one hand by internal homogeneity, and on the other hand clearly delimited from surrounding areas. For a recent survey, paying due attention to linguistic and extra-linguistic factors, see Enfield (2005). In this article, I do not take issue with earlier work arguing that Mainland Southeast Asia is a linguistic area. Rather, I want to clarify some aspects of what it means to say that Mainland Southeast Asia is a linguistic area, by measuring Mainland Southeast Asia against the results of a major recent project on areal typology, the *World Atlas of Language Structures* (Haspelmath et al. 2005), hereafter: *WALS*.

One problem with much of the earlier work on Mainland Southeast Asia as a linguistic area, as of work on the identification of language areas in general, is that it has proceeded on a largely intuitive basis, without regard for more stringent testing. For instance, often impressive lists are assembled of features that are shared by languages of Mainland Southeast Asia, but without providing the other sides of the comparison: How many languages of Mainland Southeast Asia fail to show the feature in question? How many languages outside Mainland Southeast Asia do show the feature in question? To what extent are exceptional languages in either direction on the periphery of Mainland Southeast Asia? How do languages of Mainland Southeast Asia behave with respect to features other than those included in the list?

In this article, I attempt a partial answer to the kinds of questions and concerns raised in the preceding paragraph. First, the set of

features considered is limited to those included in WALs. This means that the choice of features cannot be stacked a priori in favor of features that are shared by languages of Mainland Southeast Asia. Indeed, some of the features that are often cited as characteristic of this area, such as prevalence of monosyllabic (sometimes extended by sesquisyllabic) word structure or widespread occurrence of adversity passives, are not covered by WALs. In other words, the set of features covered by WALs was determined by criteria other than attempting to define Mainland Southeast Asia as a linguistic area, and thus avoids circularity. In all, WALs includes 142 maps, though for various reasons not all of these maps correspond to separate features that can be taken into account in the project envisaged. Three of the maps deal with correlations among features, and therefore do not depict features even logically independent of the individual features whose combinations they represent. Two of the maps relate to features of sign languages and do not provide comparability with the other maps.<sup>2</sup> Finally, one map relates to writing systems, i.e. does not relate to a structural property of language, and is thus again not comparable. This leaves in principle a total of 136 maps.

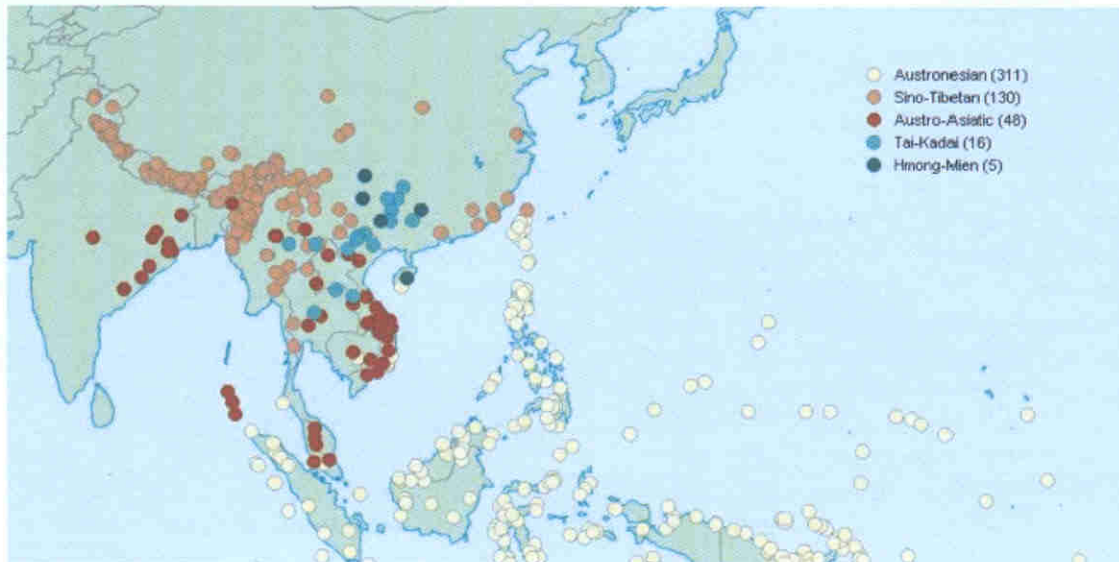
However, not all of these 136 maps are logically independent of one another. For instance, one map deals with the morphological structure of words, with

languages of Mainland Southeast Asia being characterized, as expected, as generally isolating (Dryer 2005k). But this means that several other maps, dealing with the marking of particular morphological categories by means of inflectional morphology, necessarily also show languages of Mainland Southeast Asia as being isolating, and therefore do not provide any new, independent information. In the body of this article, I consider 21 logically independent features that are represented on WALs maps, a fair subset of the approximately 100 logically independent features that are mapped in WALs with sufficient representation of languages of Mainland Southeast Asia for these to be compared with the rest of the world, and in particular neighboring areas.

It should be noted, however, that logical independence of features is not the only problem involved. At least since the pioneering work of Greenberg (1966), it has been known that there are correlations between certain logically independent features, such as a tendency for Verb–Object order to correlate with Prepositions, for Object–Verb order to correlate with Postpositions. In some cases, and sometimes indeed at least in part as a result of the possibility of testing such correlations by means of the Interactive Reference Tool that accompanies WALs, we have a good idea about the validity of correlations. But there may well be other instances where the requisite work has simply not been carried out, i.e. that features that appear to us to be independent might in fact turn out to be correlated. Eight of the features considered in this article, namely those discussed in section 2, all relate to constituent order, where there is at least the suspicion of correlations; this, of course, reduces the number of independent features that are under consideration.

---

<sup>2</sup> These maps do, incidentally, include information on Thai Sign Language, so they could provide an initial stage of the integration of sign languages into our understanding of the areal typology of Mainland Southeast Asia. However, the absence of data from other sign languages of Mainland Southeast Asia means that it is not yet possible to provide results from such a project.



Map 1: Language families of Mainland Southeast Asia and neighboring regions

In defining Mainland Southeast Asia as a linguistic area, an important factor in assessing its internal homogeneity is that the languages of the area belong to different language families, i.e. the similarities among the languages cannot be attributed solely to a common genealogical origin.<sup>3</sup> Map 1 provides a visualization of all the

<sup>3</sup> There are, of course, several hypotheses in the air that would group some of the linguistic families of Mainland Southeast Asia into a larger macro-family, and to the extent that such macro-families are valid the argument against a genealogical explanation for the internal homogeneity of Mainland Southeast Asia is weakened. However, it is still the case that few if any linguists believe that all and only the language families represented in Mainland Southeast Asia constitute a single macro-family, so there would still be some areal homogenization of the languages in question. And differences between clearly related languages spoken inside and outside Mainland Southeast Asia would still remain.

languages of Mainland Southeast Asia and adjacent regions that are mentioned at least once in the WALS maps, divided into well-established language families by means of different colors.<sup>4</sup> The major language families of the area are, in alphabetical order: Austro-Asiatic, Hmong-Mien, Sino-Tibetan, and Tai-Kadai. To this one can add Austronesian, which is, however, much more prevalent in Insular (and Peninsular) Southeast Asia than on the Mainland, although there are also a few Austronesian

<sup>4</sup> All maps in this article were prepared using the Interactive Reference Tool that accompanies WALS. While most of the maps present essentially the same information as in the printed atlas, use of the Interactive Reference Tool has two advantages: First, the resulting maps are more easily readable on standard-format paper and do not require that large format that characterizes the printed atlas. Second, reproduction of these maps is permitted by Oxford University Press's generous policy on the use for research purposes of maps generated by the Interactive Reference Tool.

languages spoken in Mainland Southeast Asia. In claiming that Mainland Southeast Asia is a linguistic area we are thus claiming that the languages belonging to these four or five language families as spoken in Mainland Southeast Asia share features that cannot be attributed to common genealogical ancestry.

Equally important is the linguistic delimitation of Mainland Southeast Asia from neighboring areas, i.e. the strongest evidence is provided by feature values that are shared by languages within Mainland Southeast Asia and not shared by surrounding areas. Of particular importance here is the fact that some language families are also spoken outside Mainland Southeast Asia, so that we can compare languages within the same family but spoken inside and outside the area of interest; differences between languages inside versus those outside the area contribute to the delimitation of Mainland Southeast Asia from neighboring regions. Of particular importance here are the Austro-Asiatic and Sino-Tibetan families, which are well represented both within and outside Mainland Southeast Asia. The Hmong-Mien and Tai-Kadai families are also represented both within and outside the area, although the languages outside Mainland Southeast Asia are, as we will see in the body of this article, in a buffer zone separating Mainland Southeast Asia from the outside, sometimes sharing feature values with Mainland Southeast Asia, sometimes not. As already noted, Austronesian languages are spoken overwhelmingly in Insular Southeast Asia or even further afield, though some, such as Cham, are spoken in Mainland Southeast Asia and are known to have assimilated in some typological respects to Mainland Southeast Asian patterns.

In addition, greater value should be placed on feature values that are found in Mainland Southeast Asia but represent relatively rare types across the world as a whole. Clearly, if a particular feature is common across the languages of the world as a whole, then it is not surprising that it should also be common in Mainland Southeast Asia. To take a concrete example, languages of Mainland Southeast Asia overwhelmingly have decimal numeral systems, but then so do 63.8% of the languages in the sample utilized in Comrie (2005), so this is not a particularly striking similarity across languages of Mainland Southeast Asia; in this case the impression is heightened by the fact that it is also the numeral system of most neighboring languages.

The ideal feature value would thus be one that is found in every language of Mainland Southeast Asia and in no other language of the world. Needless to say, no known feature value has quite such an ideal distribution, so in what follows we will be examining features that go in the direction of homogeneity within Mainland Southeast Asia accompanied by delimitation from surrounding regions, especially if the feature value in question is rare across the world as a whole. Few feature values meet all of these criteria, but as we will see, many come close.

Some final caveats are in order before turning to the empirical material. First, I have taken all information provided in the WALs maps at face value, i.e. I have not attempted to verify the correctness of any of the data presented there. No doubt some errors have crept in, though it is unlikely that they would affect the overall thrust of the conclusions presented below. And before disputing the assignment of a particular



feature to a particular language, the user of WALS materials needs always to read the text that accompanies that map, since decisions taken on assigning particular languages to particular feature values often require elaboration, such as is given in the definitions of the feature values in these texts. Second, although WALS contributors were asked as far as possible to include languages from the genealogically reasonably unbiased 100- and 200-language samples, it was up to each individual author which other languages to include, so that the samples used by different authors can be rather different (though in nearly all instances with a shared core), and the addition of languages beyond the 200-language sample may well unbalance the resulting sample genealogically. By using the Interactive Reference Tool, which provides information not only on the number of languages having a particular feature value, but also the number of genera (where a genus is a group of languages whose genealogical unity is visible by inspection) and the number of families (on a conservative assessment, omitting proposed macro-families), the statistics can be checked, at least to some extent, for genealogical adequacy. But all claims are limited by the range of languages considered by the author of the chapter in question. Thus, where I say, for instance, that no language in Mainland Southeast Asia has a particular feature value, this should always be interpreted as “relative to the sample in question”.

And finally, it should be noted that there are other ways of interpreting the injunction to follow a more clearly justified, less intuitive approach to defining linguistic areas. Thus Dahl (MS), also basing himself on the WALS database, defines a particular measure of typological distance between languages; on this measure, the typological

distances between pairs of languages across the world for which sufficient data are given in WALS to make the exercise meaningful vary between 9.8 and 74.6, with the mean and medium being virtually indistinguishable at 41.73 and 42 respectively.<sup>5</sup> Five languages of Mainland Southeast Asia—Hmong Njua (Hmong-Mien family), Khmer, Khmu’, and Vietnamese (all Austro-Asiatic family) and Thai (Tai-Kadai family) vary pairwise from 11.4 (Thai–Vietnamese) to 22.5 (Hmong Njua–Khmu’), not significantly different from the amount of typological variation found within West Germanic (represented by Dutch, English, and German, with typological distances ranging from 9.8 for Dutch–German—the lowest typological distance between any pair of languages considered by Dahl — and 21.1 for English–German). Adding the Sino-Tibetan language Eastern Kayah Li, its typological distance from the other languages of Mainland Southeast Asia ranges from 20.6 to 25, i.e. around the distance between English and German. With a typological distance of

---

<sup>5</sup> Dahl does not exclude maps that show logically related features, so given the kinds of logical dependencies among maps discussed above the similarities among languages of Mainland Southeast Asia may be somewhat inflated by Dahl’s method. For instance, nearly all languages of Mainland Southeast Asia lack gender systems, and therefore show up as having no gender not only on the map asking how many genders a language has (Corbett 2005a), but also on the two maps asking what kind of gender system a language has (Corbett 2005b; Corbett 2005c). With respect to this particular phenomenon, however, it should be noted that one of the languages considered by Dahl, namely Khmu’, is the only language of Mainland Southeast Asia in Corbett’s sample to show gender, and thus differs from all other languages of Mainland Southeast Asia considered by Dahl on all three of the gender maps in WALS.

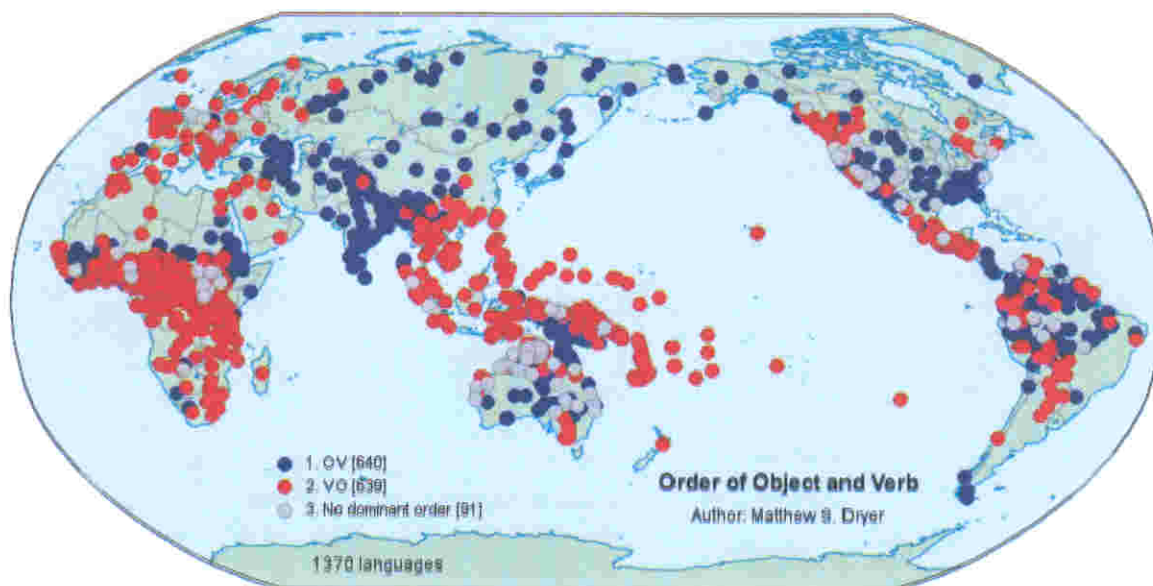
12.3, Thai and Khmer stand closer than Polish and Russian (12.8), both members of the Slavic branch of the Indo-European family. Although following a completely different methodology from that used in the present article, Dahl also concludes that Mainland Southeast Asia constitutes a clear, if not indeed the world's clearest, example of a linguistic area.

### Word order

As noted in section 1, word order typology, or more accurately: constituent order typology, has a venerable history in recent work on linguistic typology, and is represented in WALS by no fewer than 14 chapters (excluding three that deal with correlations between features). Constituent order is, moreover, one of the sets of features that most clearly sets Mainland Southeast Asia off from the rest of Asia. Constituent order does, however, bring with it a problem, namely that there are known

to be strong correlations across many of the constituent order parameters, a factor that will be given due consideration in what follows.

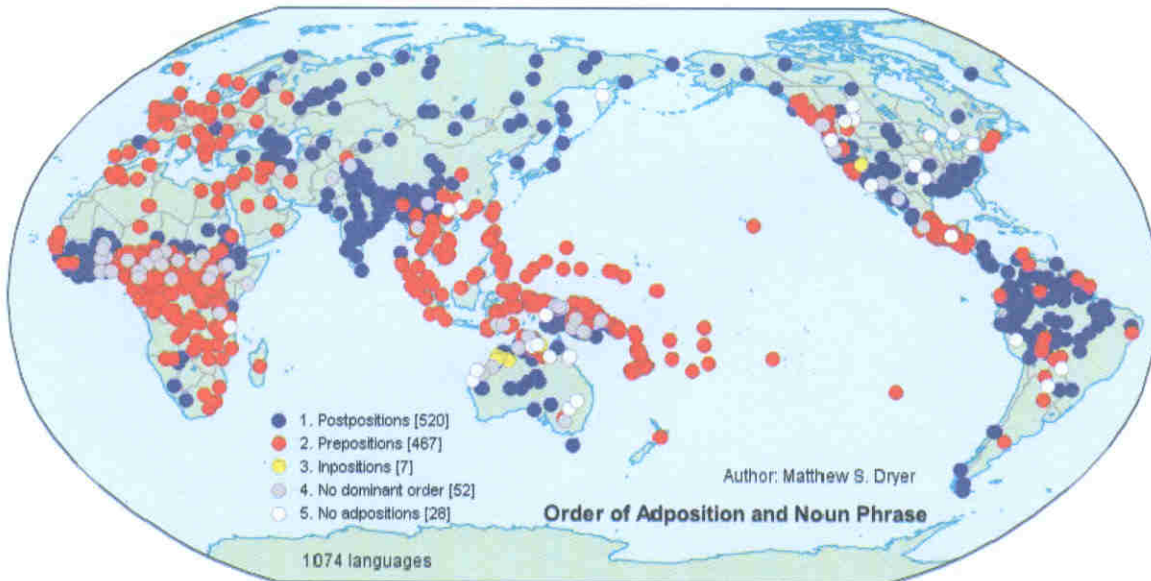
Map 2 (Dryer 2005h) shows the order of Object and Verb across the languages of the world. Dryer recognizes three values: Verb–Object (as in English ... *bought the book*), Object–Verb (as in Japanese ... *hon o katta*), and "no dominant order", the last for languages where both orders occur with no clear predominance of one over the other. Just by inspection, there is a clear difference between Mainland Southeast Asia and the rest of Asia. Asia is overwhelmingly Object–Verb, while Mainland Southeast Asia is overwhelmingly, indeed almost exclusively, Verb–Object. Across the world as a whole, both orders are equally frequent to a very high degree of approximation, so the near-solid block of Mainland Southeast Asia versus the near-solid block of the rest of



Map 2: Order of Object and Verb

Asia is highly suggestive. It will, however, be noted that in Asia the Verb–Object type does extend somewhat beyond Mainland Southeast Asia, most obviously in the case of the various varieties of Chinese. This will be important in considering other constituent order features. In the other direction, however, this feature does not distinguish Mainland Southeast Asia from Insular Southeast Asia, which is also almost exclusively Verb–Object, this will be a recurrent pattern with constituent order features. Incidentally, one might wonder whether the order of Subject and Verb (Dryer 2005j) would not serve to delimit Mainland and Insular Southeast Asia, given the frequency of Verb–Subject order among Austronesian languages; in fact, this is not the case (and thus this map is not reproduced here), since the Austronesian languages bordering Mainland Southeast Asia show a mix of Subject–Verb and Verb–Subject ordering, plus some with no dominant order.

Map 3 (Dryer 2005c) shows the order of Adposition relative to Noun phrase, with the two major types being Preposition (as in English *in the house*) and Postposition (as in Japanese *uti ni*). There are again languages with no dominant order, plus two rarer types that are not particularly significant given our areal interest, namely languages lacking adpositions—though two Hmong-Mien languages, Mien and Punu, are thus classified by Dryer—and languages with inpositions (i.e. adpositions that occur internal to their noun phrase). As with map 2, there is a clear difference between Mainland Southeast Asia and the rest of Asia, but no dividing line between Mainland Southeast Asia and Insular Southeast Asia. However, before considering map 3 as additional evidence that divides Mainland Southeast Asia from the rest of Asia, one needs to ask to what extent the feature values illustrated in maps 2 and 3 correlate. In particular, given that Mainland Southeast Asia is overwhelmingly Verb–Object, what is



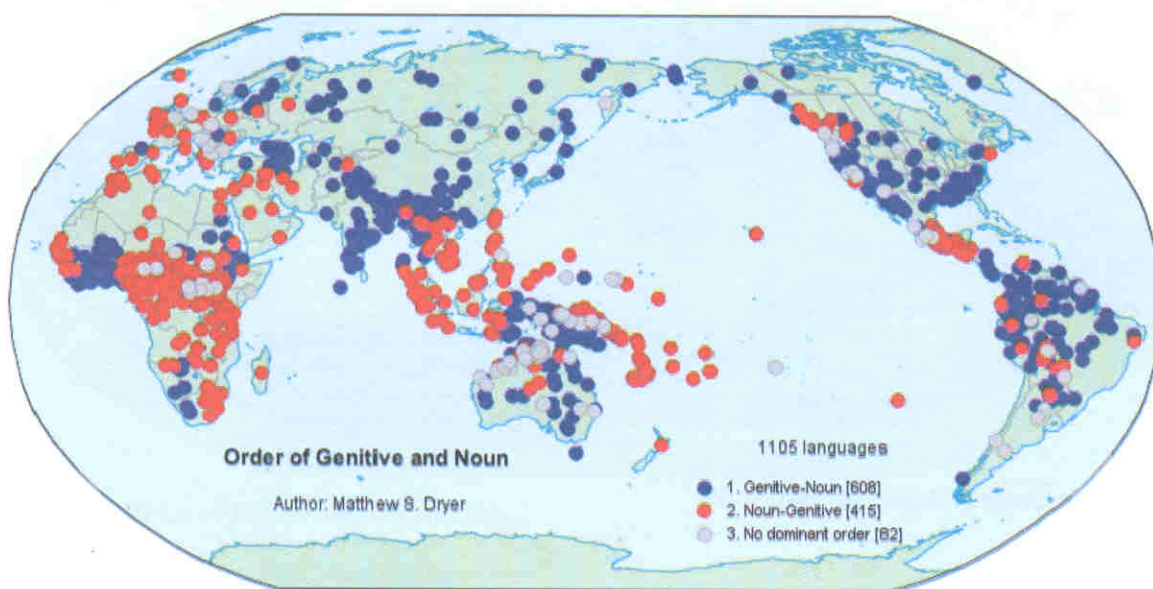
Map 3: Order of Adposition and Noun phrase



the probability of its also being Preposition–Noun phrase on the basis of a cross-linguistic valid correlation? The question can be answered by consulting Dryer (2005m). In fact, it turns out that only 8.4% of languages with Verb–Object order have Postposition, none of them in Mainland (or, indeed, Insular) Southeast Asia. Thus, map 3 provides corroborating rather than new evidence.

The next map (Map 4) shows the order of Genitive and Noun within the noun phrase (Dryer 2005g). There are two basic possibilities, Noun–Genitive (as in English *the roof of the house*) and Genitive–Noun (as in English *John's book*) — note that since English has both possibilities, with neither overwhelmingly dominant, English is in fact classified as having no dominant order. Once again, a general pattern emerges with Mainland Southeast Asia (here Noun–Genitive) contrasting with the

rest of Asia, but patterning along with Insular Southeast Asia. However, comparison of map 4 with map 2 or 3 shows that the isogloss separating Mainland Southeast Asia from the rest of Asia is different in the two cases, with “linguistic” Mainland Southeast Asia being noticeably smaller here than in the case of the order of Verb and Object or of Adposition and Noun phrase. In particular, in map 4 the order Noun–Genitive barely extends into southern China and certainly does not include varieties of Chinese, while Genitive–Noun extends into Mainland Southeast Asia, not only in the case of national languages like Burmese but even more so with smaller languages like Mlabri. This illustrates a general point that will recur in the following maps: while core Mainland Southeast Asia may differ from the rest of Asia, the dividing line will be different for different features, reflecting the fact that the present-day



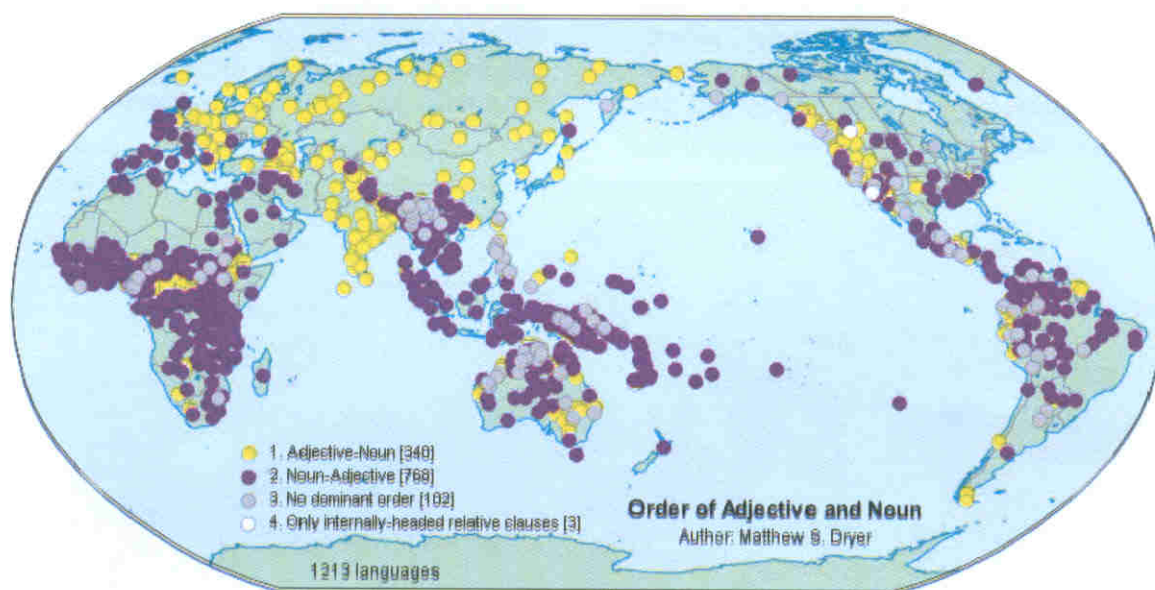
Map 4: Order of Genitive and Noun



distribution of isoglosses separating feature values is the result of complex historical processes that have not always come to a stop at the same point.<sup>6</sup>

The order of Adjective and Noun within the noun phrase (Dryer 2005b), as illustrated in Map 5, shows a similar pattern, although care is needed with the precise interpretation. There are two main constituent order possibilities, Noun-Adjective (as in Thai *bâan yây*) and Adjective-Noun (as in English *big house*), with a number of languages having no dominant order, as well as a handful, all in North America, having only internally headed relative clauses as translation

equivalents of attributive adjectives. The basic pattern seems similar to the preceding maps, with Mainland Southeast Asia rather clearly delimited from the rest of Asia, but having the same feature value as adjacent Insular Southeast Asia. However, there are some differences. The isogloss to the north runs between those for order of Verb and Object (and of Adposition and Noun phrase) and for order of Genitive and Noun. Chinese goes linguistically with the rest of Asia, while most other languages of southern China go linguistically with Mainland Southeast Asia; the order Noun-Adjective also extends well into the Himalayas.



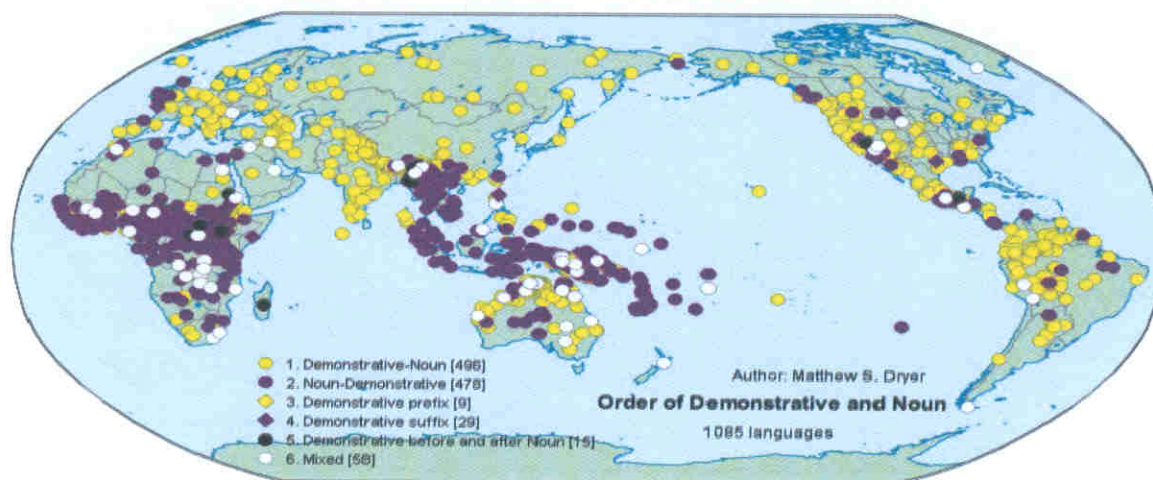
Map 5: Order of Adjective and Noun

<sup>6</sup> Readers may have been surprised that I did not include a definition of Mainland Southeast Asia early in this article. One reason for not doing so is this shifting linguistic divide.

In addition, there is an area to the northwest of Mainland Southeast Asia that might be considered transitional in that it contains a number of languages with no dominant order, the grey dots on map 5.<sup>7</sup> But there is another factor that needs to be taken into account, namely that while Verb–Object languages tend strongly to be Noun–Adjective (with English as one of the just under 20% of exceptional Verb–Object languages in this respect), Object–Verb languages are in fact somewhat more likely to be Noun–Adjective (58.8%) than Adjective–Noun (Dryer 2005l; these figures exclude languages that have no dominant order for one or the other feature). This has the result that languages combining Verb–Object and Noun–Adjective like those of Mainland Southeast Asia, which would be expected on the basis of chance to constitute 25% of the world’s languages with some dominant order for both Verb and Object and Adjective and Noun, in fact make up a staggering 40.7%.

The high frequency of this type cross-linguistically makes it less valuable as an areal diagnostic. Nonetheless, there is a clear division between Mainland Southeast Asia and most of the rest of Asia (along with much of northern and eastern Europe), which is the only large area to be characterized by the combination of Object–Verb and Adjective–Noun. So although this combination is rather rare cross-linguistically, it is dominant in Asia, and has not spread to Mainland Southeast Asia. Order of Adjective and Noun is therefore a valid isogloss separating Mainland Southeast Asia from the rest of Asia, albeit perhaps not so strikingly as some of the others.

We can now move relatively quickly through the other constituent order patterns, all of which illustrate similar patterns to those already observed, though with different dividing lines between Mainland Southeast Asia and the rest of Asia. Map 6 (Dryer 2005e) shows the



Map 6: Order of Demonstrative and Noun

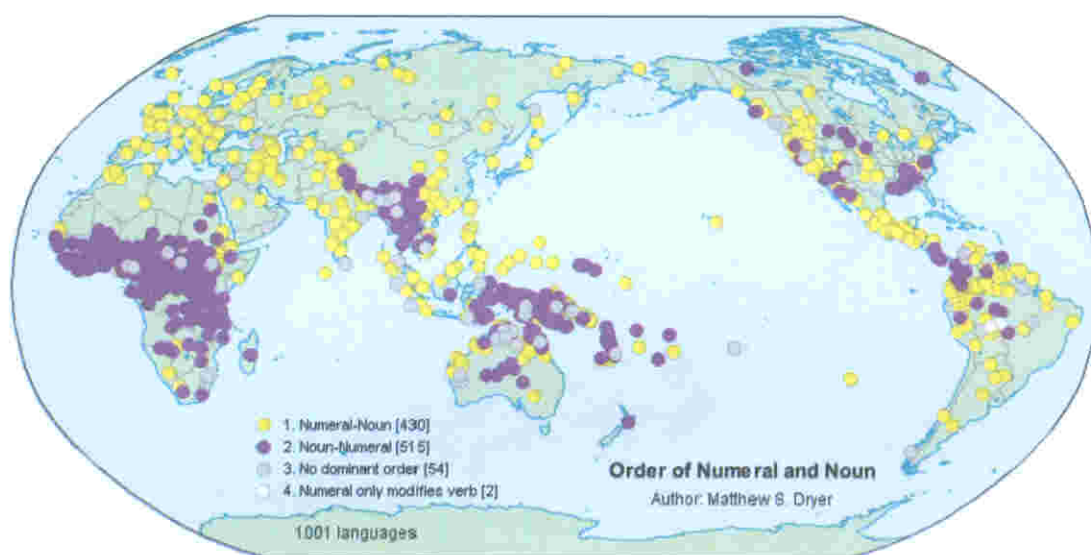
<sup>7</sup> The map as drawn here gives the impression that there is a solid “grey” area, but in fact this is an artifact of the way the software places dots in areas containing a large number of languages. In fact, this area contains a mixture of grey and purple dots.



order of Demonstrative and Noun within the noun phrase, with Demonstrative either preceding Noun (as in English *this child*) or following (as in Thai *lûuk khon nii*). Of interest here is the region to the northwest of Mainland Southeast Asia, which includes languages of a cross-linguistically rare type where Demonstrative simultaneously both precedes and follows Noun, as well as the 'mixed' type, i.e. languages that have more than one of the otherwise identified types, this region constitutes a transition area.

Map 7 (Dryer 2005g) shows the order of Numeral and Noun within the noun phrase, with the main possibilities being Numeral–Noun (as in English *ten children*)

and Noun–Numeral (as in Thai *lûuk sip khon*), as well as languages with no dominant order and a rare type, found only in South America, whereby numerals are excluded from noun phrases. The path of the isogloss is significantly different from that shown in the other constituent order maps in two respects. First, easternmost Mainland Southeast Asia goes with the bulk of Asia rather than with the rest of Mainland Southeast Asia. Second, this feature equally delimits Mainland Southeast Asia from Insular (and Peninsular) Southeast Asia. Indeed, Mainland Southeast Asia shows up here as an island surrounded by the rest of continental Asia and Insular/Peninsular Southeast Asia.<sup>8</sup>



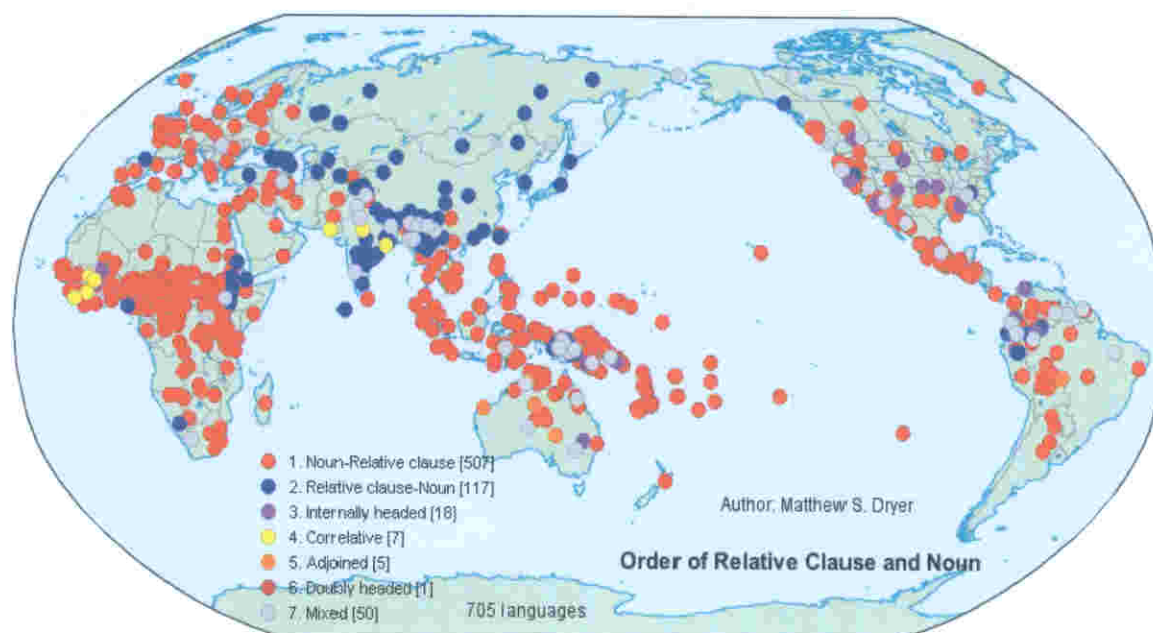
Map 7: Order of Numeral and Noun

<sup>8</sup> Although Noun–Numeral outnumbers Numeral–Noun by a ratio of about 6 : 5 in Dryer (2005g)'s sample of languages, the high number of Noun–Numeral languages is largely due to the near-exclusive occurrence of this word order in sub-Saharan Africa, to a somewhat lesser extent New Guinea, both areas with a large number of languages. Mainland Southeast Asia is the only other large area with this feature value.



Dryer (2005i) examines the order of Relative clause and Noun within the noun phrase, with the results as displayed in map 8. The major types are Noun–Relative clause (as in English *the student [who bought the book]*) and Relative clause–Noun (as in Japanese [*hon o katta*] *gakusei*), with a small number of languages representing other types (none of them in Mainland Southeast Asia) or co-occurrence of more than one type in the same language. Mainland Southeast Asia is again distinct from the bulk of Asia but goes along with Insular Southeast Asia. The dividing line is somewhat similar to that found in map 4 for Genitive and Noun, with some inroad of Relative clause–Noun into northernmost Mainland Southeast Asia.

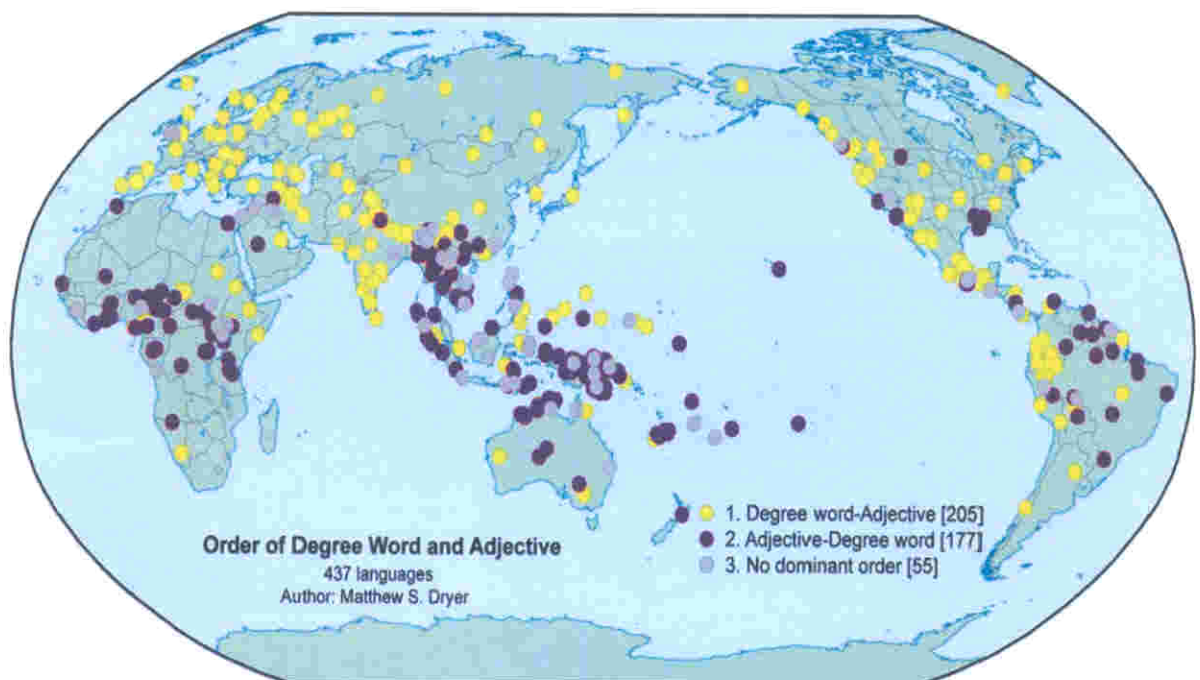
This is perhaps particularly surprising in that across the world as a whole Relative clause–Noun is by far the rarer of the two main types, outnumbered almost 5 : 1 by Noun–Relative clause. Indeed, the bulk of Asia (but here with the exception of southwestern Asia and parts of South Asia) is the only really large area with near-exclusive occurrence of Relative clause–Noun (in addition to smaller areas in interior New Guinea and a small part of East Africa). A genealogical factor may be at work here: The Relative clause–Noun languages, and also the languages with no dominant order, in northern Mainland Southeast Asia are overwhelmingly Sino-Tibetan.



Map 8: Order of Relative clause and Noun

Finally, with respect to constituent order, Map 9 shows the order of Degree word and Adjective (Dryer 2005d), with the possibilities being Degree word–Adjective (e.g. English *very good*), Adjective–Degree word (e.g. Thai *dii nák*), or dominance of neither. The separation of Mainland

Southeast Asia from the rest of Asia is clear, even if there is some extension of the “no dominant order” type into Mainland Southeast Asia. The boundary to Insular Southeast Asia is much less clear, since all three types are found in the neighborhood, despite the fact that these are all Austronesian languages.

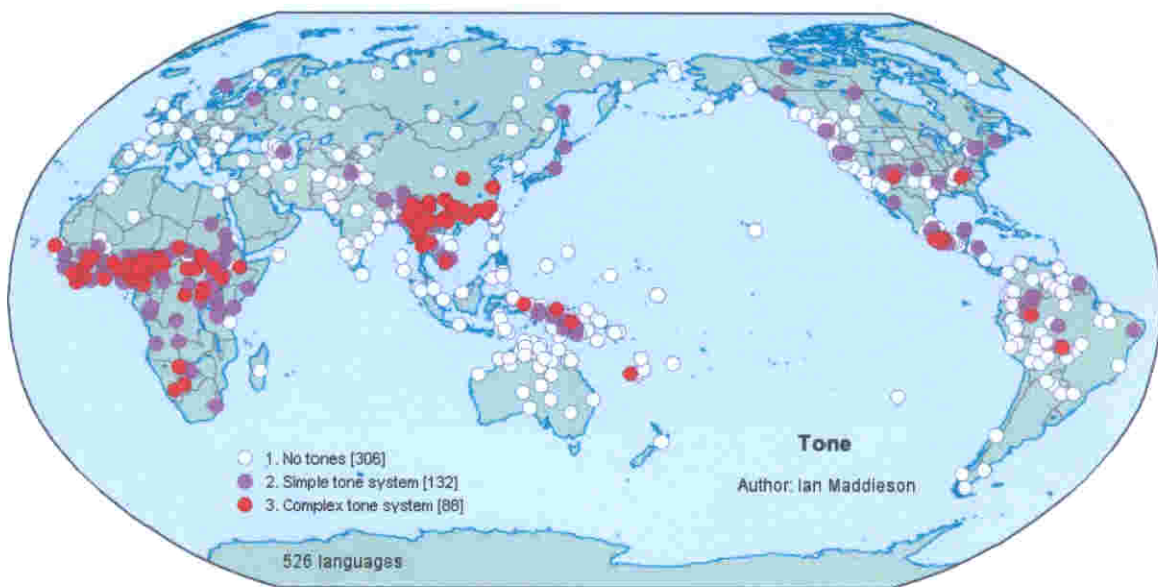


Map 9: Order of Degree word and Adjective

## Phonology

Turning now to phonology, pride of place inevitably goes to tone (Maddieson 2005c). As shown in map 10, Maddieson divides the languages of the world into three types, those with a simple tone system (i.e. a binary tonal opposition), those with a complex tone system (three or more phonemically opposed tones), and those without phonemic tone. It is clear from the map that Mainland Southeast Asia is overwhelmingly characterized by complex tone systems, a pattern that extends north into Chinese. The only other part of the world of comparable size with a similar incidence of complex tone systems is equatorial Africa, though here the distribution is more patchy. In Mainland

Southeast Asia there is also some occurrence of simple tone systems, including on the northwest boundary towards the rest of Asia. The rest of Asia is almost exclusively non-tonal, with occasional simple tone systems. This time, the boundary to Insular Southeast Asia is even more clear, since this area completely lacks tone systems. It should be noted that some of the minority of non-tonal languages shown in Mainland Southeast Asia have other prosodic features, such as voice register in Austro-Asiatic languages, that are known to be closely linked historically to tone, so that with a slightly different definition of features Mainland Southeast Asia might have been even more homogeneous than it appears in map 10.

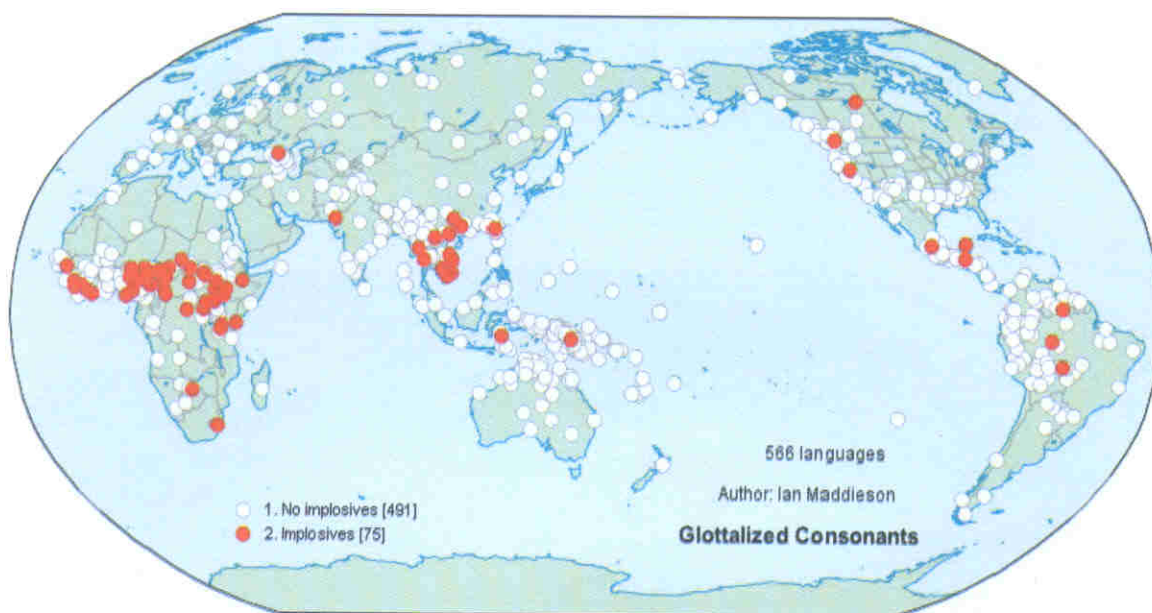


Map 10: Tone



Map 11 is based on Maddieson (2005b), but is simplified to show only the distinction between languages that have implosive consonants and those that do not.<sup>9</sup> The map shows that implosives are indeed thick on the ground in Mainland Southeast Asia, although there are also a number of languages in the area that lack them. (Again, the software that generates the map somewhat distorts the frequencies, giving preference to red over white dots.) So in this

map, Mainland Southeast Asia is not so homogeneous as in the others. However, given the rarity of implosives across the languages of the world — they are found in only 13.3% — the fact that they are so frequent in Mainland Southeast Asia, and virtually lacking in surrounding areas, is a significant fact of areal distribution. The only other part of the world with a similar incidence of implosives is equatorial Africa.<sup>10</sup>



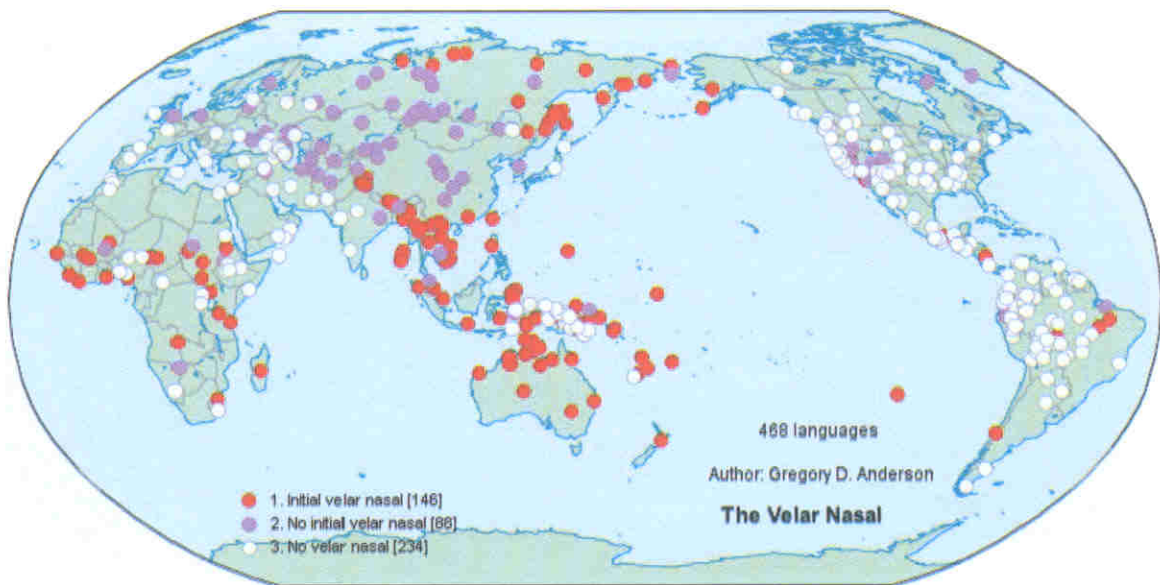
Map 11: Implosives

<sup>9</sup> The original map includes other kinds of stops with a glottalic airstream mechanism, namely ejectives and glottalized resonants. The former are not found in Mainland Southeast Asia, the latter only in languages of the area that also have implosives.

<sup>10</sup> The reader will have noted that typological similarities between Mainland Southeast Asia and equatorial Africa recur in several maps. Further exploration of this parallelism lies outside the scope of this article.

The next phonological map, map 12 (Anderson 2005), shows the distribution of the phonemic velar nasal across the languages of the world, with three values: no velar nasal; velar nasal only in non-initial position in the word; velar nasal also in initial position. Mainland Southeast Asia is characterized by overwhelming incidence of initial velar nasals, the only exception in the sample being Khmer (which does have velar nasals in other positions). Incidence of initial velar nasals spreads into the Himalayas, but otherwise shades off into occurrence of velar nasals

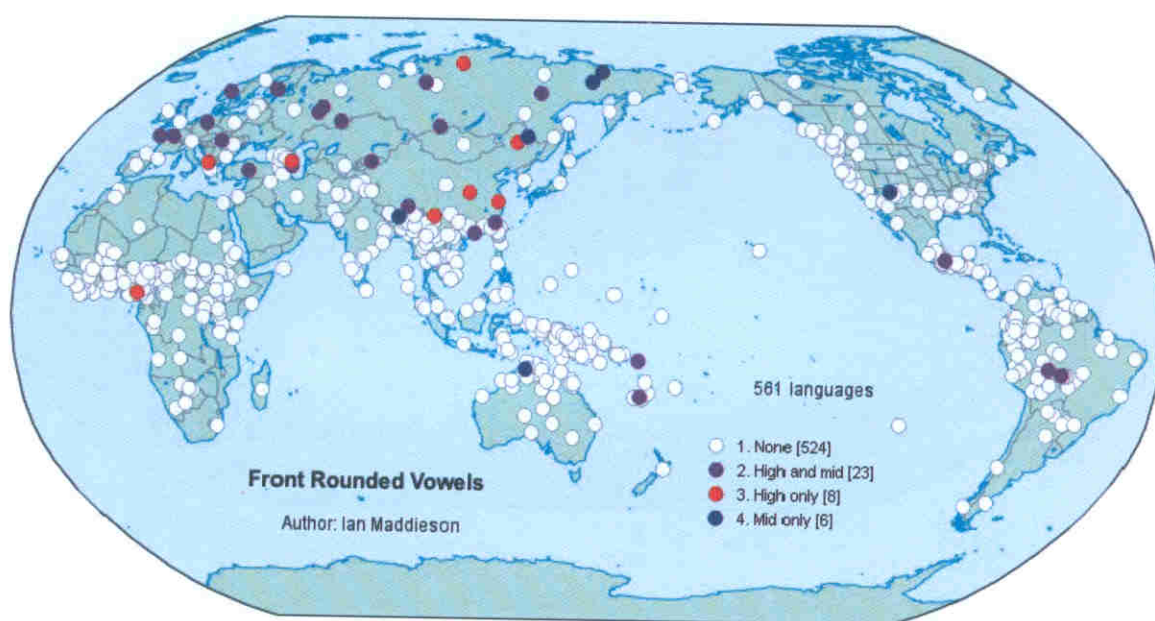
only non-initially in adjacent parts of Asia, with the isogloss separating varieties of Chinese — in Anderson's sample, Cantonese has initial velar nasals, Mandarin does not. (In northeastern Asia the possibility of an initial velar nasal picks up again, but this is presumably a separate area, given the broad swathe of intervening languages that lack this possibility.) There is, however, no boundary for this feature between Mainland and Insular Southeast Asia, since the latter also contains many languages with word-initial velar nasals.



Map 12: The velar nasal

The last phonological map, map 13 (Maddieson 2005a), shows the distribution of phonemic front rounded vowels in the languages of the world. Inclusion of this map may at first seem puzzling, since although Mainland Southeast Asia is homogeneous in the absence of such sounds, this is also by far the majority pattern in the languages of the world: in Maddieson's sample, 93.4% of the world's languages lack front rounded vowels. However, the part of the world where front vowels occur frequently cross-linguistically is clearly identifiable as northern Eurasia. While to some extent the incidence of the phenomenon coincides with the occurrence of palatal vowel harmony, as in Uralic,

Turkic, and Mongolic languages, which gives a strong impetus to having front counterparts to back rounded vowels, it is also found in languages that lack palatal vowel harmony, such as the Germanic languages and varieties of Chinese. Moreover, the area spreads almost to the boundary of Mainland Southeast Asia — the last outpost shown on the map is Cantonese — but resolutely refuses to go further. So even though this is not so clear a case as the others discussed in this article, there is still some evidence for a linguistic barrier between Mainland Southeast Asia and the rest of Asia. (Insular Southeast Asia is like Mainland Southeast Asia in lacking front rounded vowels.)



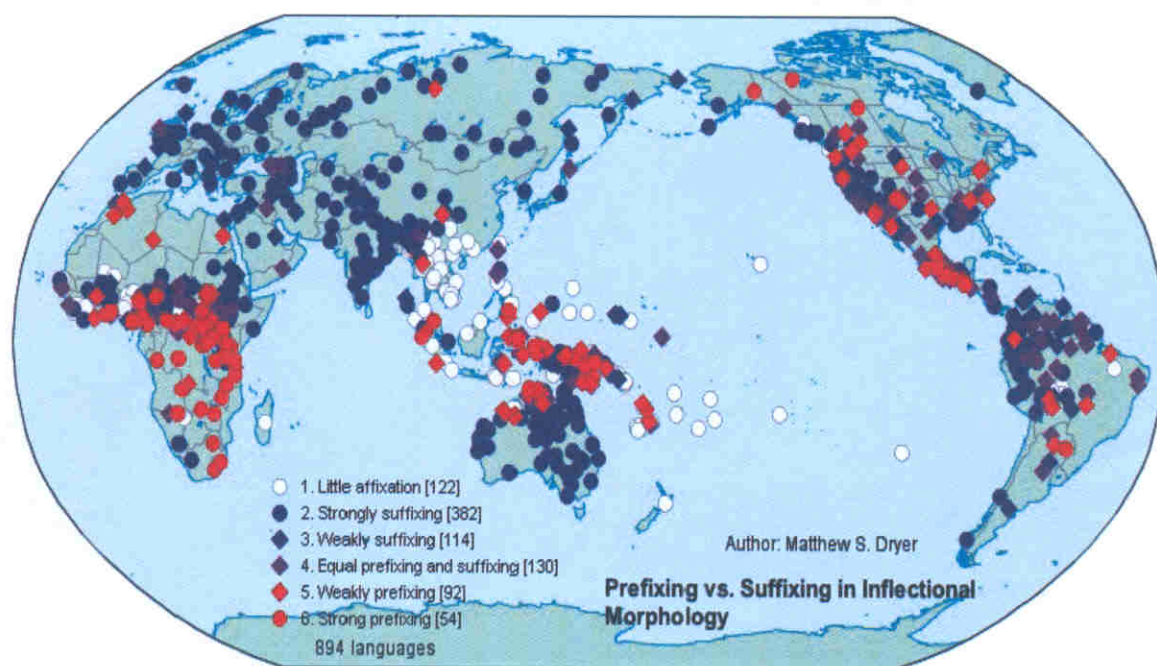
Map 13: Front rounded vowels



## **Morphology and morphological categories**

One of the often noted properties of languages of Mainland Southeast Asia is their isolating typology, with virtually no inflectional morphological categories and even relatively little use of bound morphemes in derivational morphology. Map 14 (Dryer 2005k) illustrates this with respect to inflectional morphology. In this map, the white dots represent languages with little or no affixation for inflectional

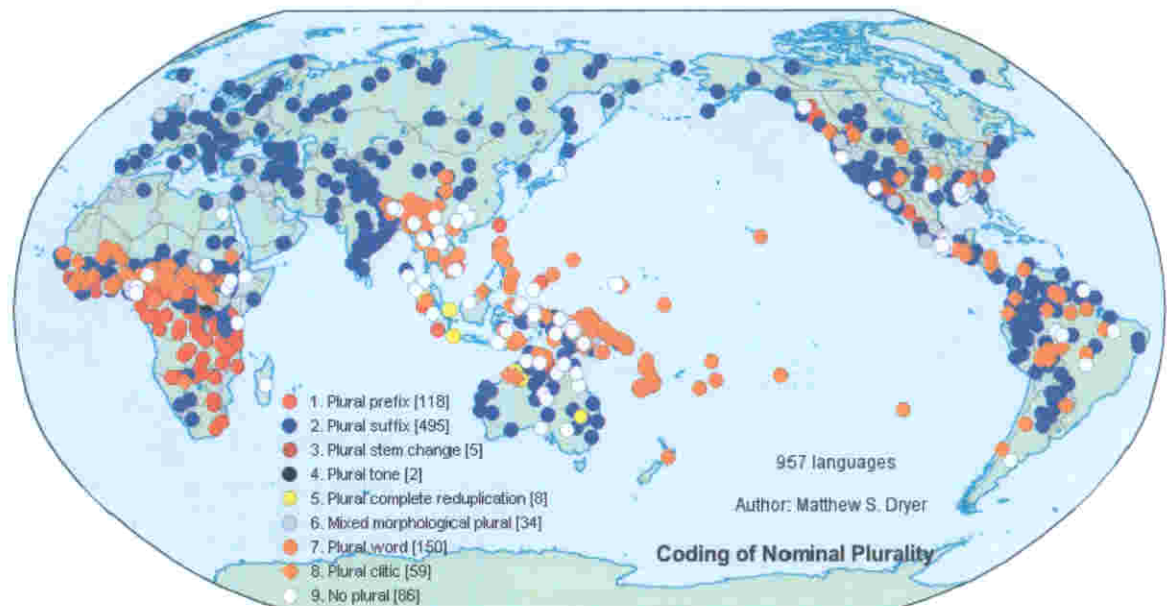
morphology, and Mainland Southeast Asia stands out as the largest area that consistently shows only white dots. Indeed this feature value extends somewhat beyond Mainland Southeast Asia to the north and west, although to the west there is much geographical intermixing with suffixing languages, which represent the majority of the world's languages and also the near-universal pattern in the rest of Eurasia. There is also some extension to Insular Southeast Asia, though here mixed geographically with both suffixing and prefixing languages.



Map 14: Prefixing versus suffixing in inflectional morphology

Given that languages of Mainland Southeast Asia show little or no inflectional morphology, it might seem irrelevant to go on to ask whether they encode nominal plurality, and if so how. But as map 15 (Dryer 2005a) shows, languages can code nominal plurality not only through affixal or other morphology, even if this is by far the most usual means cross-linguistically, but also by means of plural words, as in Vietnamese *những chó* 'dogs', literally 'PLURAL dog'. Map 15 shows that while Mainland Southeast Asia does contain a fair number of languages lacking plural marking, it also contains a number of languages, including Khmer and Vietnamese, with plural words. The orange circles indicating plural words and the orange lozenges indicating plural

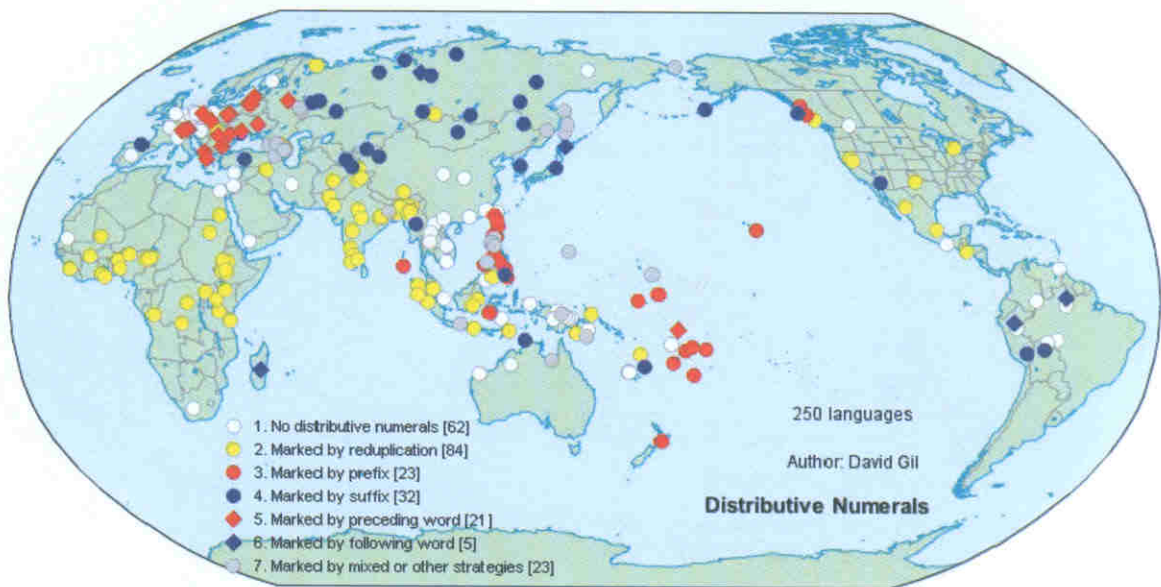
clitics extend well to the north and west, though in the northwest there are some languages that have plural suffixes, not shown clearly in the map resolution used here. But the rest of Eurasia is almost exclusively characterized by plural suffixes. The boundary to Insular Southeast Asia is less clear, since in regions adjacent to Mainland Southeast Asia one finds languages with no plural and with plural words, as in Mainland Southeast Asia, but also languages using prefixes and reduplication to mark plurality. It should be noted that in Dryer's sample only one language of Mainland Southeast Asia, the Austro-Asiatic language Sre, has a plural affix, namely a prefix.



Map 15: Coding of nominal plurality

A distributive numeral is a special numeral form with meaning ‘*n* each’ for a numeral *n*. English and the languages of Mainland Southeast Asia are alike in lacking such special forms, although many other languages do have such forms, e.g. German *je drei* ‘three each’, where *drei* is the numeral ‘three’ and *je* is a preceding word dedicated to marking distributive numerals. While it might seem surprising to speakers of English or Thai to consider lack of distributive numerals a striking feature of a language, in fact map 16 (Gil 2005a) shows that this is the minority possibility across the world’s languages, with just under 25% of the world’s languages having no dedicated distributive numerals. In this sense, then, Mainland Southeast Asia is again a linguistic area,

though this time extending well to the north to take in both Chinese (including Mandarin) and the Turkic language Salar, and contrasting clearly with northern Asia and South Asia, both of which have dedicated distributional numerals, usually marked by suffixes in the former area, by reduplication in the latter. The boundary to the east is clear-cut, with Austronesian languages of Taiwan and the Philippines solidly having distributional numerals, usually marked by means of prefixes, although there is more fluidity at the boundary to the rest of Insular Southeast Asia, where languages lacking distributive numerals are found alongside those that have them (with the latter usually using reduplication).

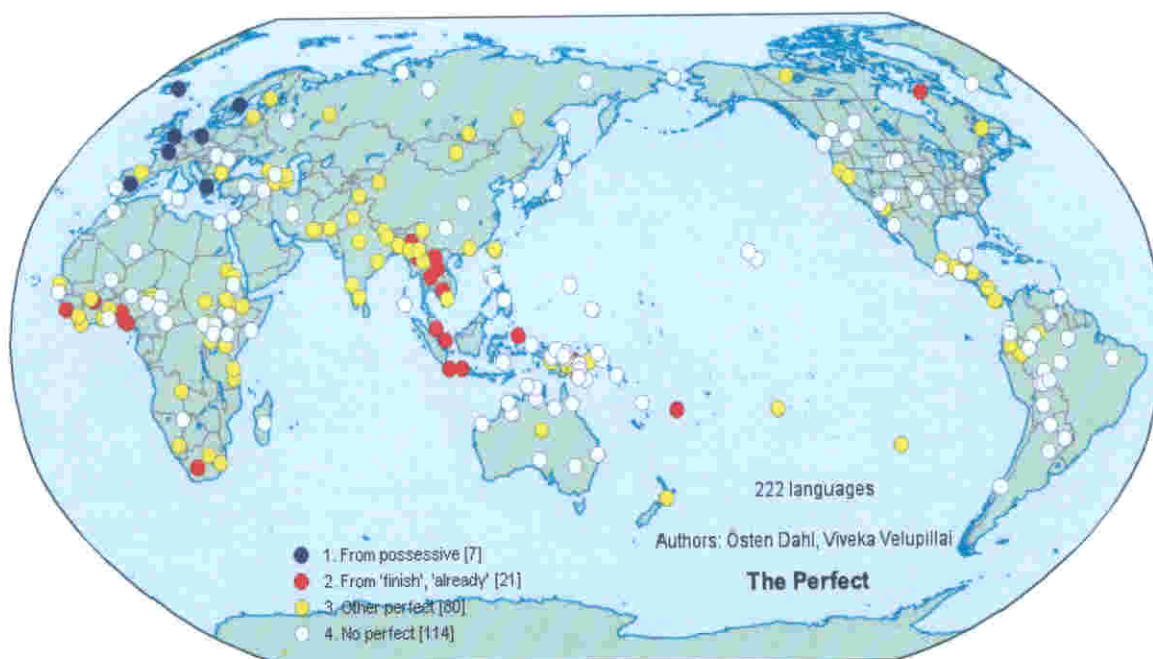


Map 16: Distributive numerals



Most of the maps referring to tense–aspect categories are restricted to inflectional marking, so predictably languages of Mainland Southeast Asia lack such inflectional categories by virtue of lacking inflectional morphology. However, one of the maps (Dahl and Velupillai 2005), dealing with the Perfect, includes also non-inflectional means of marking the Perfect, in particular possessive-derived constructions (as in English *we have eaten*) and those using a word meaning ‘finish’ or ‘already’ (e.g. *lěew* in Thai *raw kin khâaw lěew*). As map 17 shows, this latter construction, indicated by red dots, is widespread in Mainland Southeast Asia, spreading somewhat to the northwest and

especially to the south, namely Peninsular Malaysia and western Indonesia. The type is otherwise rare in the world, making up only 9.5% of the languages in Dahl and Velupillai’s sample, 19.4% of those with a Perfect. Most languages with a Perfect have what Dahl and Velupillai call ‘other Perfect’, which of course subsumes all possibilities other than possessive-derived and ‘finish’/‘already’, so that the 36% of the world’s languages that belong to this type, or the 74% of the world’s languages that have a Perfect, almost certainly do not constitute a single type. There are some such languages, including Vietnamese, in Mainland Southeast Asia.

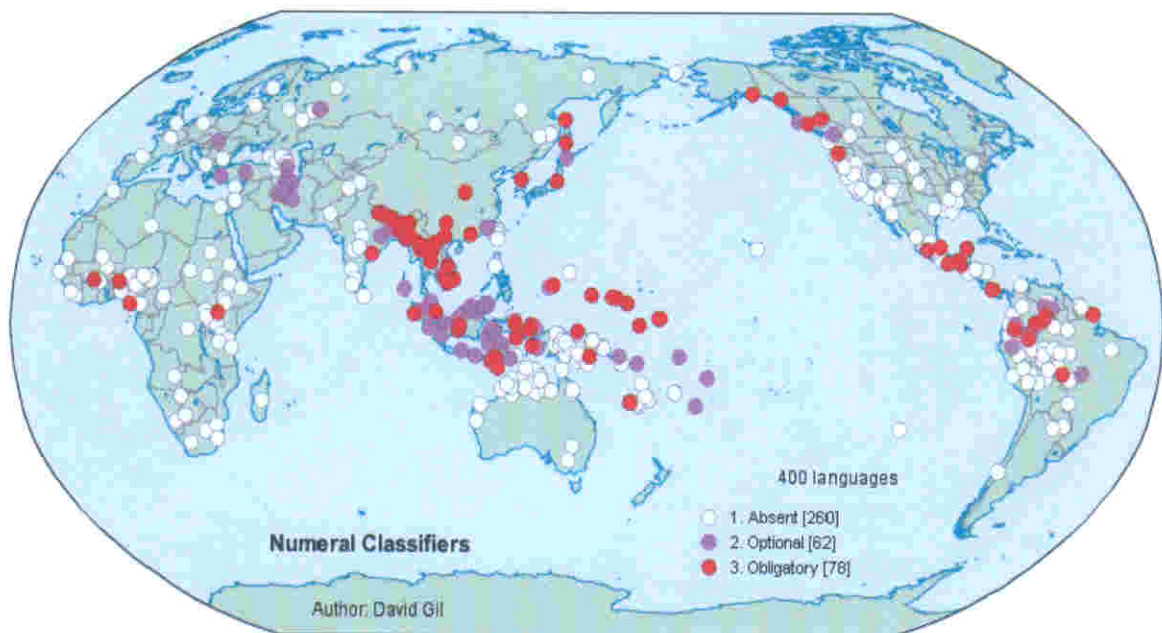


Map 17: The Perfect

### Other features

Numeral classifiers, such as *khon* in Thai *lûuk sip khon* 'ten children', are a well-known feature of languages of Mainland Southeast Asia, and map 18 (Gil 2005b) clarifies the sense in which this holds. First, Mainland Southeast Asia belongs to an area where, with very few exceptions, the use of numeral classifiers is obligatory. This includes all languages in Gil's sample except for Khmer, where their use is optional, but also extends well beyond Mainland Southeast Asia to the northwest and especially to the north, running through Chinese, Korean, and Japanese up to Nivkh on Sakhalin island and the mouth

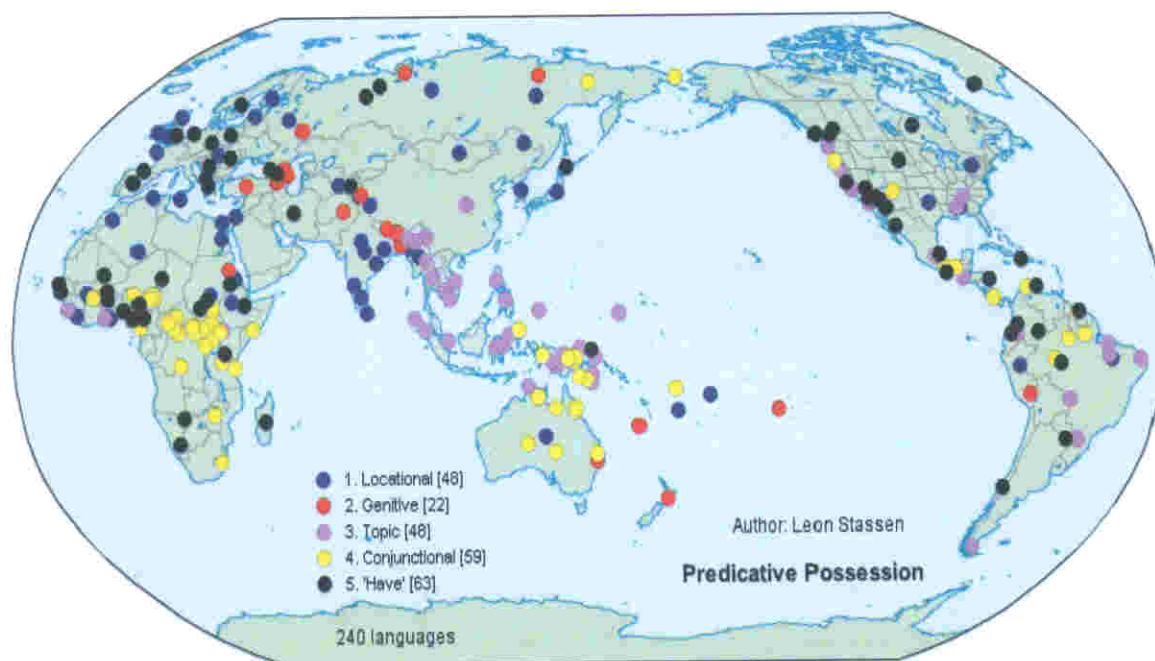
of the Amur River; in Gil's sample, only the Chinese variety Hokkien and Ainu stand apart in having only optional use of classifiers. The boundary to the northwest is marked by a small area of intermingling with languages having optional classifiers followed by absence of classifiers, the norm in the rest of Asia. Moving towards Insular Southeast Asia, one encounters an area where most languages have optional classifiers, a few none or obligatory classifiers, in Peninsular Malaysia and western Indonesia; the Austronesian languages of Taiwan and the Philippines lack numeral classifiers, thus providing an abrupt boundary to the east.



Map 18: Numeral classifiers

Map 19 (Stassen 2005c) shows that the languages of the world vary considerably in their translation equivalents of the English construction 'you have a house', the English construction here reflecting the 'have' type. The type that interests us particularly in map 19 is that represented by the pink dots, the so-called topic type, as in Thai *khun mii bâan* (lit. you be house, or, somewhat more revealingly: as for you, there is a house). This type is found in 20% of the languages in Stassen's sample, which is in one sense as close to chance as one could get given that there are five types in his typology, but nonetheless the distribution of the topic

type is striking: All languages of Mainland Southeast Asia belong to this type, which extends somewhat to the northwest and north (encompassing Mandarin Chinese), although it is absent from the rest of Eurasia. In the other direction, there is no such boundary, since the topic type is also characteristic of the western part of the Austronesian-speaking world, and is also found (though not exclusively) further east, e.g. in the New Guinea area, where it includes some non-Austronesian (so-called "Papuan") languages. Where it occurs elsewhere in the world — equatorial Africa, the Americas — it is sporadic.

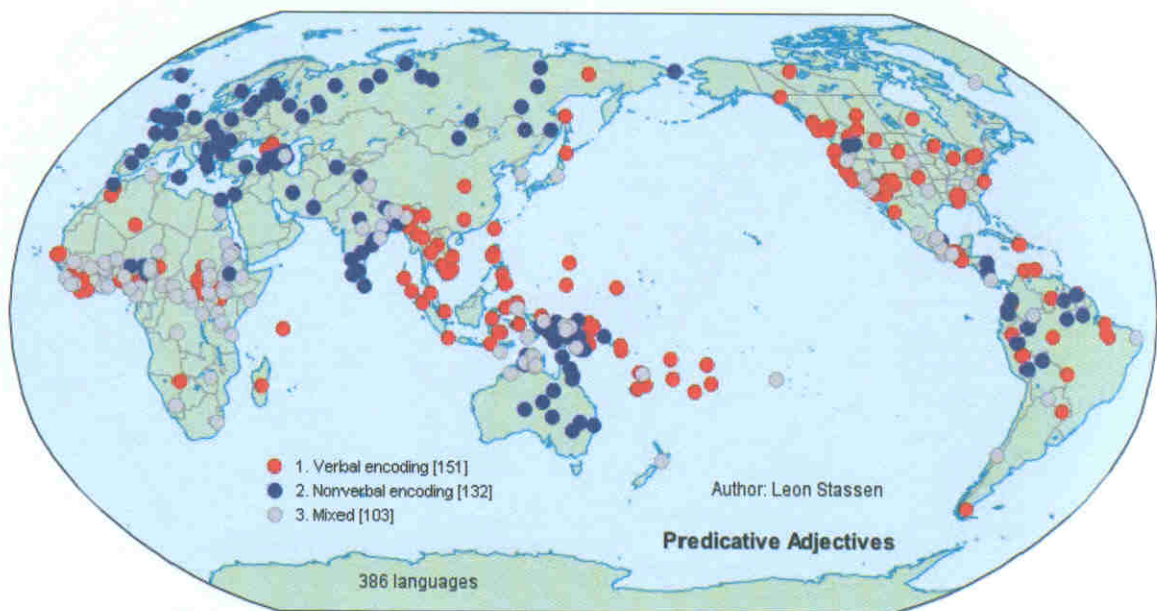


Map 19: Predicative possession



Another aspect of predication is covered by map 20 (Stassen 2005b), namely the question of whether adjectives are encoded primarily like verbs or not. Exclusively verbal encoding is found in some 39% of the languages in Stassen's sample, more than either of the other two types, but it is nonetheless striking that Mainland Southeast Asia is a solidly "verby" area, contrasting with most of Eurasia, which is equally "non-verby". With respect to this feature, Mainland Southeast Asia extends a little further to the northwest and to the north (to encompass Mandarin Chinese), with a number of languages with mixed encoding of adjectives on the northwest

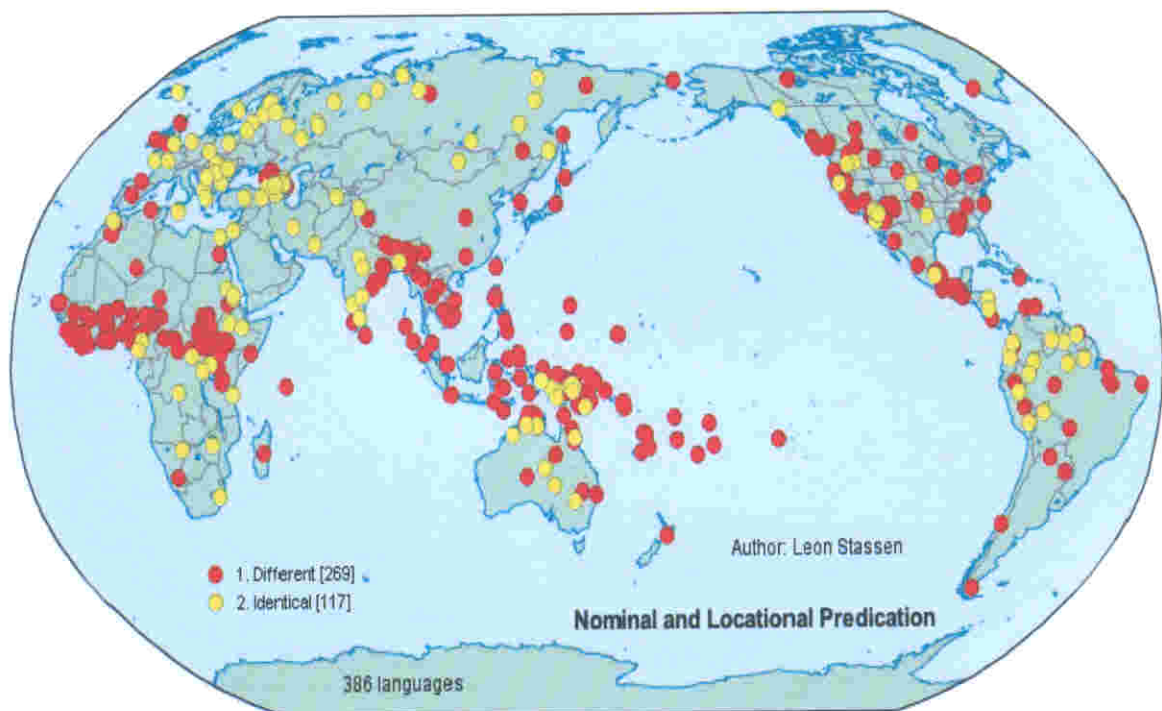
boundary. Further to the northeast there is the intriguing observation that Korean and Japanese have mixed encoding, but a line of verbal encoding then takes up again with the isolates/small families Ainu, Nivkh, and Yukaghir, though it is not clear to what extent this is a real areal phenomenon. With respect to this feature, however, there is absolutely no boundary between Mainland Southeast Asia and Insular Southeast Asia, which is equally "verby" in its encoding of adjectives; only towards New Guinea and Australia do we find a boundary, with mixed languages and then exclusively "non-verby" languages.



Map 20: Predicative adjectives

One striking difference between English and Thai is that English uses the same verb *be* to introduce both nominal and locational predicates, as in *you are a teacher* and *Malee is at the market*, whereas Thai distinguishes these two as *pen* and *yüu*, as in *khun pen khruu* and *maalii yüu thii talàat*. Thai here represents by far the majority pattern among the languages of the world, as can be seen from map 21 (Stassen 2005a). In Stassen's sample, 69.7% of the languages make the distinction. There is, nonetheless, a clear difference between Mainland Southeast

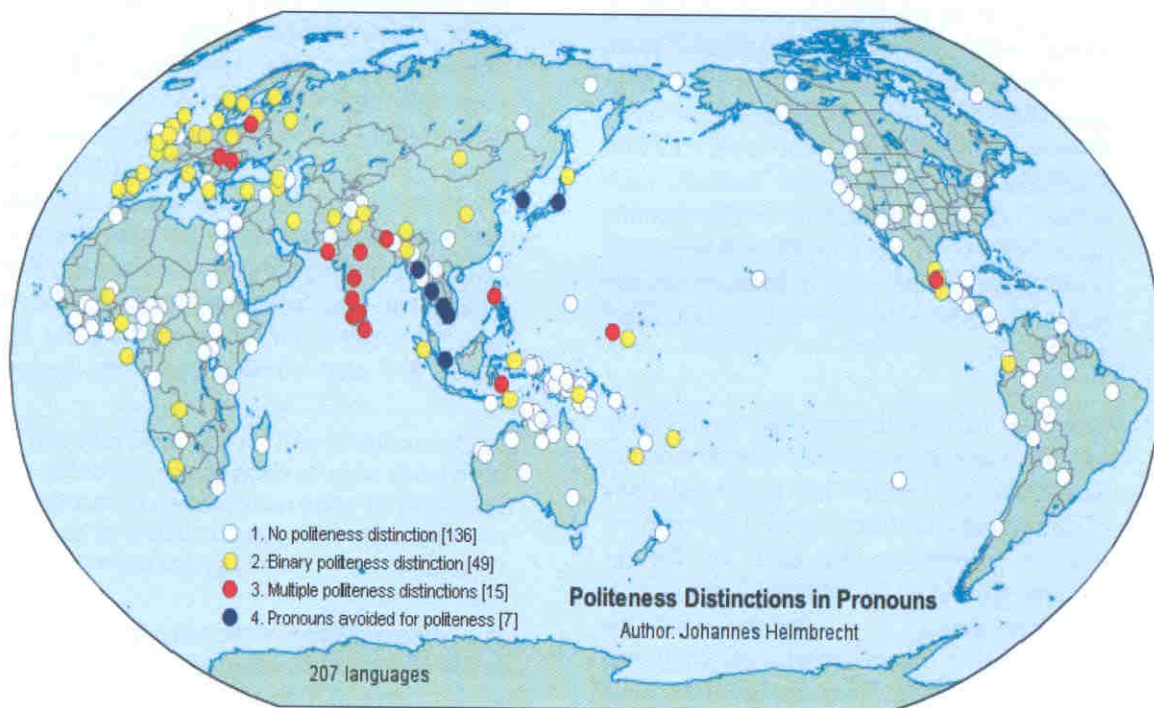
Asia and most of the rest of Eurasia, with Mainland Southeast Asia containing exclusively languages that make the distinction, whereas most of Eurasia is the single largest area of languages that do not (with only sporadic exceptions). However, the area that encompasses Mainland Southeast Asia is considerably larger for this feature than for the others discussed, first in that this time all of Insular Southeast Asia is included, and second in the extent to which this type extends to the west and north, well into eastern India and across the eastern part of northeastern Asia.



Map 21: Nominal and locational predication

While most features discussed in this article have been purely structural, the last one to be considered hovers on the boundary of structural properties and sociolinguistics, since it deals with politeness distinction in pronouns (Helmbrecht 2005). Map 22 shows that only 34.3% of the languages in Helmbrecht's sample show politeness distinctions, with the rarest way of encoding politeness being the avoidance of pronouns, found in only 7 out of 207 languages (3.4%). Three of these languages are spoken in core Mainland Southeast Asia, namely Thai, Khmer, and Vietnamese, with Burmese on the periphery of Mainland

Southeast Asia and Indonesian in Insular Southeast Asia. The remaining two languages, Korean and Japanese, are separated from Mainland Southeast Asia by China, so the extent to which we might be dealing with a single area rather than two distinct areas here would require further investigation. Nonetheless, pronoun avoidance as a politeness strategy clearly characterizes at least all the socially dominant languages of a somewhat extended Mainland Southeast Asia and sets this area apart not only from the most of the rest of Asia but indeed from most of the rest of the world.



Map 22: Politeness distinctions in pronouns



## Conclusion

In this article I have tried to show that a number of features, in general logically independent of each other, and selected on the basis of what data are provided in WALS, demonstrate that Mainland Southeast Asia is indeed an internally homogeneous linguistic area. It is most clearly set off from the rest of Asia, though with some fluidity in the boundary for different features, with such regions as the northeast of South Asia and China often showing up as transition zones, sometimes going with Mainland Southeast Asia, sometimes with the rest of Asia. There is in general no such clear dividing line between Mainland Southeast Asia and Insular Southeast Asia. Only a few features, such as tone, show a boundary at or close to this divide. Nonetheless, it is clear that overall homogeneity of the region decreases as one moves from Mainland to Insular Southeast Asia.

One way of quantifying this would be to ask, for various languages, whether or not they share the feature that is characteristic of Mainland Southeast Asia. I carried out such a calculation for each of the 21 features discussed in sections 2 through 5. (In most cases, it is obvious what the feature value characteristic of Mainland Southeast Asia is. Two cases might need clarification. For coding of plurality, I assume that either no plural marking or use of plural words or clitics counts as Mainland Southeast Asia. For numeral classifiers, only obligatory use of numeral classifiers counts as Mainland Southeast Asia.) I then tested these features against three national languages of Mainland Southeast Asia, namely Thai, Khmer, and Vietnamese, plus three languages that belong to the periphery of the region,

namely Burmese, Mandarin, and Indonesian.

On this measure, Thai has 19 of the relevant feature values (the only exceptions being the formation of nominal plurals, for which no information is given for Thai,<sup>11</sup> and implosives), Khmer 18 (the exceptions being tone, initial velar nasal, and obligatory numeral classifiers), and Vietnamese also 18, although for one feature, namely order of Adjective and Degree word,<sup>12</sup> no information is provided for Vietnamese (the two exceptions are order of Numeral and Noun and “finish/already”-type Perfect). Of the languages in the periphery, Indonesian shares 14 of these features, Burmese 11, while Mandarin shares 7 (with no information given for one feature, namely prefixing versus suffixing versus neither in inflectional morphology).<sup>13</sup> While at best suggestive — a fuller study would need to take all features into account (as Dahl MS does), and also consider not only logical

<sup>11</sup> The exclusion of Thai may have reflected uncertainty as to its classification given the possibility of using reduplication to encode plurality with a “very small number of nouns” (Smyth 2002: 25); Lao is shown as having no nominal plural.

<sup>12</sup> My interpretation of the material presented in Thompson (1965, especially p. 267) would be that the usual Vietnamese translation equivalent of ‘very’ is *rất*, which precedes the adjective, i.e. Vietnamese would in fact be an exception to the usual Mainland Southeast Asian pattern.

<sup>13</sup> My own assessment would be that Mandarin is at least mildly suffixing, differing in this respect from Cantonese, which is included in the relevant map and classified as having little affixation. In this respect, as also in having initial velar nasals, Cantonese would thus be closer to the Mainland Southeast Asian type than Mandarin.

dependencies and globally valid correlations across them, but also whether particular features need to be weighted more than others as markers of linguistic areas in general, or of Mainland Southeast Asia in particular — this does point to a core Mainland Southeast Asia area comprising roughly Thailand, Laos, Cambodia, and Vietnam, with a periphery extending through Myanmar, China, and Peninsular and Insular Southeast Asia.

Also on this measure, Thai turns out to be the most typical of the three major national languages of Mainland Southeast Asia considered here. Interestingly, although Dahl (MS) uses a completely different methodology, he also concludes that, among the languages he considers, Thai is the most typical for Southeast Asia. This is intriguing given that Tai languages are relative newcomers to Mainland Southeast Asia, moving from southwestern China into

an area that was then mainly Austro-Asiatic and Sino-Tibetan-speaking (Enfield 2005: 185). Two factors would need to be taken into account in a fuller investigation of this phenomenon: first, the extent to which consideration of the minority languages of Mainland Southeast Asia might alter this picture, with some of these perhaps being even more typically Mainland Southeast Asian than Thai; and second, the extent to which Thai might have adopted already existing typical Mainland Southeast Asian features in the course of its southward migration, through contact with the non-Tai languages spoken in the area at the time of Thai's expansion, including substrate effects from speakers of these languages being assimilated into the Thai speech community. The present study perhaps raises as many questions as it answers, but I hope that investigation of these questions will prove fruitful for our understanding of Mainland Southeast Asia, and not only from a linguistic perspective.

## References

- Anderson, Gregory D.S. 2005. 'The velar nasal (*ŋ*)', in Haspelmath et al. 2005: 42–45.
- Comrie, Bernard. 2005. 'Numeral bases', in Haspelmath et al. 2005: 530–533.
- Corbett, Greville G. 2005a. 'Number of genders', in Haspelmath et al. 2005: 126–129.
- Corbett, Greville G. 2005b. 'Sex-based and non-sex-based gender systems', in Haspelmath et al. 2005: 130–133.
- Corbett, Greville G. 2005c. 'Systems of gender assignment', in Haspelmath et al. 2005: 134–137.
- Dahl, Östen. MS. 'An exercise in a posteriori language sampling'.
- Dahl, Östen and Viveka Velupillai. 2005. 'The perfect', in Haspelmath et al. 2005: 271–272, 280–281.
- Dryer, Matthew S. 2005a. 'Coding of nominal plurality', in Haspelmath et al. 2005: 138–141.
- Dryer, Matthew S. 2005b. 'Order of adjective and noun', in Haspelmath et al. 2005: 354–357.
- Dryer, Matthew S. 2005c. 'Order of adposition and noun phrase', in Haspelmath et al. 2005: 346–349.
- Dryer, Matthew S. 2005d. 'Order of degree word and adjective', in Haspelmath et al. 2005: 370–373.
- Dryer, Matthew S. 2005e. 'Order of demonstrative and noun', in Haspelmath et al. 2005: 358–361.
- Dryer, Matthew S. 2005f. 'Order of genitive and noun', in Haspelmath et al. 2005: 350–353.
- Dryer, Matthew S. 2005g. 'Order of numeral and noun', in Haspelmath et al. 2005: 362–365.
- Dryer, Matthew S. 2005h. 'Order of object and verb', in Haspelmath et al. 2005: 338–341.
- Dryer, Matthew S. 2005i. 'Order of relative clause and noun', in Haspelmath et al. 2005: 366–369.
- Dryer, Matthew S. 2005j. 'Order of subject and verb', in Haspelmath et al. 2005: 334–337.
- Dryer, Matthew S. 2005k. 'Prefixing versus suffixing in inflectional morphology', in Haspelmath et al. 2005: 110–113.
- Dryer, Matthew S. 2005l. 'Relationship between the order of object and verb and the order of adjective and noun', in Haspelmath et al. 2005: 394–397.
- Dryer, Matthew S. 2005m. 'Relationship between the order of object and verb and the order of adposition and noun phrase', in Haspelmath et al. 2005: 386–389.
- Dryer, Matthew S. 2005n. 'Relationship between the order of object and verb and the order of relative clause and noun', in Haspelmath et al. 2005: 394–397.
- Enfield, N.J. 2005. 'Areal linguistics and Mainland Southeast Asia'. *Annual Review of Anthropology* 34: 181–206.
- Gil, David. 2005a. 'Distributive numerals', in Haspelmath et al. 2005: 222–225.
- Gil, David. 2005b. 'Numeral classifiers', in Haspelmath et al. 2005: 226–229.
- Greenberg, J.H. 1966. 'Some universals of grammar with particular reference to the order of meaningful elements', in J.H. Greenberg (ed), *Universals of Language*, 2 ed. Cambridge, MA: MIT Press, 73–113.



- Haspelmath, Martin, Matthew Dryer,  
David Gil, and Bernard Comrie,  
eds. 2005. *The World Atlas of  
Language Structures*. Oxford:  
Oxford University Press.
- Helmbrecht, Johannes. 2005. 'Politeness  
distinctions in pronouns', in  
Haspelmath et al. 2005: 186–189.
- Maddieson, Ian. 2005a. 'Front rounded  
vowels', in Haspelmath et al.  
2005: 50–53.
- Maddieson, Ian. 2005b. 'Glottalized  
consonants', in Haspelmath et al.  
2005: 34–37.
- Maddieson, Ian. 2005c. 'Tone', in  
Haspelmath et al. 2005: 58–61.
- Smyth, David. 2002. *Thai: An Essential  
Grammar*. London: Routledge.
- Stassen, Leon. 2005a. 'Nominal and  
locational predication', in  
Haspelmath et al. 2005: 482–485.
- Stassen, Leon. 2005b. 'Predicative  
adjectives', in Haspelmath et al.  
2005: 478–481.
- Stassen, Leon. 2005c. 'Predicative  
possession', in Haspelmath et al.  
2005: 474–477.
- Thompson, Laurence C. 1965. *A  
Vietnamese Grammar*. Seattle:  
University of Washington Press.