DE GRUYTER MOUTON Linguistics Vanguard 2022; 8(s1): 11–23

Leah Pappas* and Gary Holton

A quantitative approach to sociotopography in Austronesian languages

https://doi.org/10.1515/lingvan-2020-0044
Received May 21, 2020; accepted May 10, 2021

Abstract: Absolute spatial orientation systems are pervasive and diverse among Austronesian languages, and decades of research has suggested that such systems are motivated at least in part by environmental and cultural factors. In this paper, we take a quantitative approach to the study of orientation systems by presenting the results of an exploratory multifactorial analysis of spatial orientation systems across 131 Austronesian languages, representing nearly all available data on orientation systems for the family. We analyze these data using multinomial logistic regression to uncover correlations between orientation type and four predictor variables representing cultural and environmental factors: geographic distribution, economy, geography (proximity to the sea), and ruggedness of terrain. Our model suggests that while not entirely predictive of the type of orientation system, the factors geography and economy alone account for much of the variation among spatial orientation systems in our sample, supporting a "weak" form of the Sociotopographic Model (Palmer, Bill, Jonathon Lum, Jonathan Schlossberg & Alice Gaby. 2017. How does the environment shape spatial language? Evidence for sociotopography. *Linguistic Typology* 21(3). 457–491). Additionally, this study demonstrates the potential of quantitative analytical methods for exploring the relationship between culture, environment, and spatial orientation systems.

Keywords: Austronesian; sociotopography; spatial orientation

1 Introduction

Spatial orientation has been the subject of extensive study across Austronesian languages due to the prevalence of geocentric (absolute) systems of spatial orientation in which spatial coordinate systems are anchored in local or global geography rather than intrinsic properties of the reference object or the viewer.[1] While relative and intrinsic frames may play varying roles in these languages, "referential systems operating within the absolute frame appear to be universal in Austronesian languages" (Palmer 2002: 111).

The basic principle of Austronesian spatial orientation has long been assumed to be based on a distinction between landward (Proto-Austronesian [PAN] *daya) and seaward (PAN *lahud) (Adelaar 1997: 43, Blust 2013: 311). However, a recent survey of orientation systems in the Malayo-Polynesian languages outside Oceanic reveals a great diversity of orientation systems, particularly within East Nusantara. Here, straddling the Wallace Line, we find systems of orientation employing axes anchored by a number of different environmental features, including rivers, coastlines, elevation, the path of the sun, and wind. In addition, many languages employ hybrid systems which combine aspects of more than one of these types. Languages employing a simple land–sea opposition are cqsurprisingly rare in this region (Holton and Pappas to appear).

1 We use the term "geocentric" to refer to absolute systems (in the sense of Levinson 2003) which are anchored in the environment, as well as to landmark-based and geomorphic systems (in the sense of Bohnemeyer et al. 2015).

***Corresponding author: Leah Pappas,** University of Hawai'i at Mānoa, Honolulu, Hawaii, USA, E-mail: lpappas@hawaii.edu. https://orcid.org/0000-0002-3329-7686
Gary Holton, University of Hawai'i at Mānoa, Honolulu, Hawaii, USA, E-mail: holton@hawaii.edu. https://orcid.org/0000-0002-9346-1572

Though there are significant areal patterns, the distribution of spatial orientation systems cross-cuts linguistic classifications, inviting an extralinguistic explanation. Among the many variables, culture and geography have been repeatedly proposed as potential external influences on the structure (and usage) of orientation systems. This is particularly salient from a historical perspective, as it can be difficult to find language-internal motivations for the evolution of diverse spatial orientation systems. Surveying the grammaticization of locatives across a sample of 104 Oceanic languages, Bowden (1992: 131) notes the "ability for culture and geography to determine what can be linguistically significant." Beyond linguistics, the relation between social status and spatial orientation has figured prominently in anthropological research. Bubandt describes the "upcoast" region in Buli as a domain which is "both socially and morally distinct from the rest of social space" (1997: 147). This leads Senft to suggest that spatial systems transcend linguistic knowledge through the important role of cultural, historical, and geographical knowledge in organizing space and that description of spatial language can only be completed using "interdisciplinary anthropological linguistic approaches" (1997: 32).

Given that geocentric systems of spatial orientation are by definition anchored in geography, it is not surprising that geographic factors are frequently asserted to explain the distribution of orientation systems. Moreover, it is well known that vocabulary often "bears the stamp of the physical environment in which the speakers are placed" (Sapir 1912: 229). Thus, we find rich terminologies relating to riverine orientation in regions where rivers provide major points of geographic reference but fewer such terms in coastal regions where rivers are less prominent features. Beyond these simple observations, some authors have made much stronger claims, asserting the existence of a correlation between type of orientation system and local geography. This notion has been most clearly articulated in the Topographic Correspondence Hypothesis (Palmer 2015), which proposes that absolute spatial systems are both anchored in the surrounding geography as well as motivated by that geography. The Sociotopographic Model (Palmer et al. 2017) refines this further by proposing that sociocultural factors mediate between spatial systems and geography. Evidence in support of the Sociotopographic Model has been accumulated from a variety of languages in differing geographic locations (cf. Palmer 2015; Palmer et al. 2017; Heegård and Liljegren 2018; Schlossberg 2018), supporting the notions that (a) orientation systems may be shared across genetically unrelated languages, and (b) genetically related languages spoken in differing geographic environments may have radically different orientation systems.

However, despite the importance of extralinguistic factors in explaining spatial orientation typology, it is clear that neither culture nor geography fully determines the choice of orientation system. Rather, spatial cognition is the result of the "complex interplay of language structure, local environment, cultural practices, and language use" (Palmer et al. 2017: 488). Hence, it is relatively easy to find counterexamples to the Sociotopographic Model. This is of course unsurprising: We don't expect speakers to change orientation systems immediately upon moving to a new environment, so apparent "mismatches" between local geography and type of orientation system will naturally arise.[2] But such exceptions do not invalidate sociotopographic approaches. As we show below, while cultural and geographic factors do not completely determine orientation systems, significant correlations do exist between at least some geographic and cultural factors and the choice of orientation system.

While numerous studies have illustrated these correlations with regional or local case studies, in this paper we take a typological approach to understanding the relationship between culture, geography, and spatial orientation. Rather than using experimental methods to uncover spatial orientation types in particular languages, we draw on data presented in existing grammatical descriptions in a large sample of Austronesian languages. For each language in the sample, we identified spatial orientation type based on existing grammatical descriptions. We then additionally coded each language for four social and geographic variables: the geographic distribution of the language, the main economy of its speakers, the proximity of the speakers to the sea, and the ruggedness of the terrain in the area.[3] We then developed an exploratory statistical model which

2 However, orientation systems can indeed change as the result of population migrations to new geographic locations; see Holton (2017).

3 Details on the coding methodology can be found in the Supplementary Materials at https://doi.org/10.5281/zenodo.4708029.

predicts the type of orientation system based on these social and environmental variables (Section 2). The results of this analysis are discussed in Section 3 and summarized in Section 4.

2 Methodology

Drawing on extant literature—which included published and unpublished grammatical descriptions, spatial descriptions, and dictionaries—we coded 131 Malayo-Polynesian languages according to the type of spatial orientation system employed: land–sea, coastal, elevation, riverine, or cardinal. Systems in which the land–sea opposition is indistinguishable from geophysical elevation were coded separately (labeled land–sea*). Further, languages which combine aspects of more than one type of system—for example, a land–sea axis combined with a cardinal axis, as in Balinese (Wassmann and Dasen 1998)—were coded as having both a primary and a secondary orientation system. We label these "hybrid" systems.

Given the relatively small size of our language sample, it was necessary to limit the number of classes in the spatial orientation typology in order to ensure model fit. We thus take two approaches to coding languages with hybrid systems in order to achieve an overall reduction in the number of orientation systems in the typology. In the first approach, which we call the "combined" method, we combine primary and secondary axes into a hybrid system type, yielding three additional classifications: riverine + cardinal, land–sea + elevation, and land–sea + cardinal. These three hybrid systems represent the most common types of combinations of orientation systems in our sample. Using a second method, which we call the "doubled" method, we treat languages with hybrid systems as two separate languages that only vary in the type of system, while all other coding remains the same. The distribution of languages by orientation type using the combined method is shown in Figure 1.

For each language we also coded four sociotopographic (social and environmental) factors: distribution, economy, geography, and terrain (see Table 1). Each factor was treated as a categorical variable with two or three values. Distribution refers to the geographic spread of a language and may range from a few villages or cities (village distribution) to an entire island (island distribution) or across several islands (distributive distribution). Economy represents the primary economic activity of a language group: subsistence economies, communities that produce one or more cash crops (agricultural economy), or those that are primarily neither subsistence nor agricultural (diversified economy). Geography refers to a language's proximity to the ocean; languages may be spoken along the coast (coastal geography), in interior regions (non-coastal geography), or in a combination of both environments (diversified geography). Finally, terrain is a measure of the relative variation in altitude in the region where the language is spoken; the terrain may be relatively flat (non-mountainous terrain) or quite rugged (mountainous terrain).

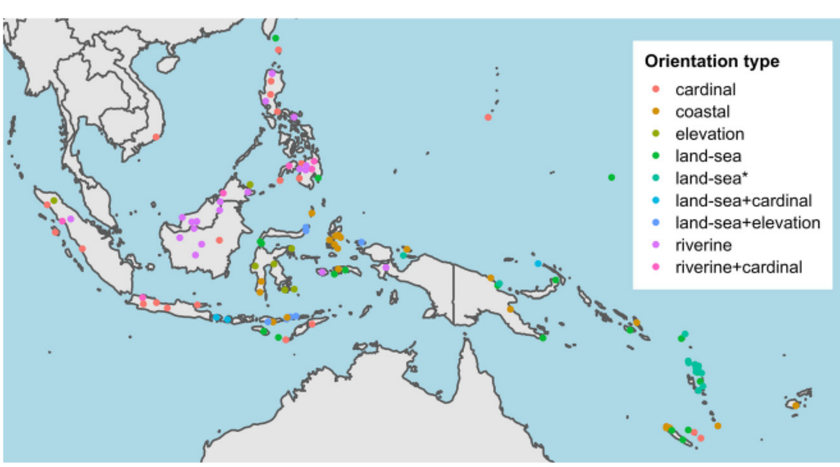

Figure 1: Location of the 131 languages in the sample according to orientation type (combined method). Language coordinates are from Glottolog (Hammarström et al. 2020). The map was produced using Leaflet for R (Graul 2016).

Table 1: Coding of sociotopographic factors.

Factor	Values	Description
Distribution	village; island; distributed	geographic spread of a language
Economy	subsistence; agricultural; diversified	primary economic activity
Geography	coastal; inland; diversified	proximity to ocean
Terrain	mountainous; non-mountainous	topographic relief

We model the relationship between orientation type and sociotopographic factors using a statistical model which predicts orientation type using the four sociotopographic factors as independent variables (predictors). Since orientation type is an unordered categorical variable and the independent variables are polytomous, we employed logistic regression, implemented using the *multinom()* function from the nnet package in R (Venables and Ripley 2002). Initial attempts to train the model using subsets of the data introduced large sampling errors, possibly due to the relatively small size of the data set and the large number orientation types. Some subsets completely failed to represent some types of orientation systems with low-frequency in the data set. Therefore, we trained the model using the entire data set of orientation type codings, using both the "combined" and "doubled" approaches. This approach does not allow us to directly assess the accuracy of the model; however, given the exploratory nature of this work, we do not see this as a great disadvantage.

The models report the probability of a language having a particular type of orientation system as opposed to some chosen reference orientation type, for a given set of values for the four predictor variables.[4] We ran the models with land–sea as the reference orientation type, given that (a) a land–sea system is reconstructed for Proto-Austronesian, and (b) land–sea is the most frequent type of system in our data. Alternatively, we could have used cardinal as the reference orientation type, as it is the second most frequent system in the data and also plays a major role in many of the languages of wider communication used throughout the region. However, setting cardinal as the reference type resulted in no significant difference in model fit, so we retained land–sea as the reference type. The odds reported for our model thus reflect the likelihood of a given type of orientation system occurring rather than a land–sea system.

Given the large number of predictors and the potential for autocorrelation, we attempted to improve model fit by selectively removing predictors. Comparing Akaike's Information Criterion (AIC) values indicates that the best fit is achieved using only the economy and geography predictors (see Table 6 in the Appendix). Although the improvement in model performance is slight (with an AIC of 347.93 for the combined approach, rather than an AIC of 363.48 using all predictors), this simplified model has the additional advantage of capturing sociocultural features and geographic-environmental features using just two independent variables.[5]

Tables 2 and 3 report the log odds (ratios) for the combined and doubled approaches, respectively, using the economy and geography predictors. The models fit fairly well, accounting for roughly 50% of the variability in the data (McFadden's R^2 is 0.495 for the combined approach and 0.454 for the doubled approach). The columns in the tables correspond to type of orientation system; land–sea is not represented here because it is set as the reference system. The rows in the tables correspond to values of the predictors as compared to a default reference value for the predictor. Choice of reference value for the predictors does not affect the model fit but does change the way we describe the model. We set reference values arbitrarily as diversified for the factors of geography and economy (and also, in other models, distributed for distribution, and non-mountainous for terrain). Log odds in the table can thus be interpreted as the likelihood of a particular orientation system (shown in the column heading) rather than land–sea, for a given value of the predictor variable (shown in the row heading) as opposed to the reference value for that predictor. Thus, for example, the large value for cardinal and inland shown in Table 3 implies that cardinal systems (rather than land–sea) are more likely with inland as opposed to diversified geography. Since log odds is the natural log of odds ratio,

4 In multinomial logistic regression, the reference type (or reference category) is the baseline type to which the likelihood of other types occurring instead of that type is measured. This is generally chosen to be the most normative or expected type.

5 For more details on model comparison, see https://gmholton.github.io/mp-space/sociotopography.html.

Table 2: Log odds (combined method, economy and geography predictors only).

	Orientation type							
	Cardinal	Coastal	Elevation	Land–sea*	Riverine	Land–sea + cardinal	Land–sea + elevation	Riverine + cardinal
Agriculture	−1.36	−1.40	−1.50	−1.32	−0.35	−0.37	−6.69†††	−38.23†††
	(0.25)	(0.35)	(0.43)	(0.39)	(0.84)	(0.83)	(0.00)	(0.00)
Subsistence	−2.23†	0.28	−1.75	0.50	−0.80	−24.77††	14.03†††	−3.09†
	(0.04)	(0.79)	(0.29)	(0.66)	(0.61)	(0.00)	(0.00)	(0.06)
Coast	−1.75	−0.50	16.01†††	15.78†††	−2.88†	−2.37	14.75†††	18.96†††
	(0.15)	(0.72)	(0.00)	(0.00)	(0.09)	(0.17)	(0.00)	(0.00)
Inland	−14.88†††	−12.39†††	34.69†††	4.36†††	18.54†††	−16.00	−0.51†††	39.4†††
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.99)	(0.00)	(0.00)
Intercept	2.84††	0.75	−16.56†††	−15.49†††	0.95	1.13	−29.44†††	−18.88†††
	(0.01)	(0.57)	(0.00)	(0.00)	(0.49)	(0.4)	(0.00)	(0.00)

p-values in parentheses. Note that many of the significant relationships are among hybrid systems, which had much smaller sample sizes than the non-hybrid systems. AIC: 347.93, McFadden's R^2: 0.495. †$p < 0.1$. ††$p < 0.05$. †††$p < 0.01$.

negative values of log odds correspond to odds ratios less than 1:1. That is, where log odds is negative, land–sea systems are more likely than the given system.

Because we were not able to successfully train the model on a subset of the data, we cannot easily assess the accuracy of the model. In other words, while the model fits well to our data set, we cannot assess the ability of the model to predict orientation systems for novel data (i.e., languages not included in our sample), given relevant values of the predictors. This shortcoming aside, we do not expect such a model to have a high degree of accuracy. For one, our model makes use of just a few simplistic social and geographic variables and thus will fail to capture the full complexity of sociotopographic influences on the choice of orientation system. But perhaps more significantly, even if the Sociotopographic Model holds, we do not expect social and topographic factors to be the *only* drivers of orientation system choice. Linguistic factors—particularly structural and historical factors—clearly play an important (and perhaps even primary) role, as likely do other nonlinguistic factors not considered here. Thus, our use of the term "model" should not be taken to mean that we are attempting to represent all of the potential factors which drive the choice of orientation system. Rather, our approach might be better described as a "partial model," in that we attempt to model only some of the many potential factors behind orientation system type. Nevertheless, the fact that this model can account for *some* of the variation suggests that the factors considered here are relevant, at least for the data considered in this sample.

As will be discussed, using the combined and doubled approaches results in two slightly different models. Since the doubled approach counts languages with a hybrid orientation system twice, the underlying data set for the doubled model is much larger ($N = 169$). In order to reduce the number of values for orientation type coding, the combined approach required many of the low-frequency hybrid systems to be ignored. The

Table 3: Log odds (doubled method, economy and geography predictors only).

	Orientation type				
	Cardinal	Coastal	Elevation	Land–sea*	Riverine
Agriculture	1.32	0.83	0.31	0.73	0.37
	(0.15)	(0.54)	(0.85)	(0.61)	(0.76)
Subsistence	−1.41	1.09	0.02	1.03	−1.68
	(0.14)	(0.36)	(0.99)	(0.39)	(0.16)
Coast	1.11	0.20	−14.52†††	−15.01	0.97
	(0.18)	(0.86)	(0.00)	(1.00)	(0.42)
Inland	16.26†††	−3.48†††	16.28†††	−3.67†††	19.35†††
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Intercept	−0.49	−1.41	−1.60	−1.25	−1.41
	(0.53)	(0.22)	(0.18)	(0.27)	(0.16)

p-values in parentheses. AIC: 368.84, McFadden's R^2: 0.454. †$p < 0.1$. ††$p < 0.05$. †††$p < 0.01$.

doubled approach thus has the advantage that it does not require any categories to be discarded and instead allows us to capture the full diversity of orientation systems across the languages. However, in practice we found little difference between the models. In the following section we discuss the results in more detail, drawing on both of the models.

3 Results and discussion

In this section we interpret the results of the models, referencing the statistically significant correlations identified in the two-factor models (Tables 2 and 3). We first discuss the geography predictor (Section 3.1), followed by the economy predictor (Section 3.2). We then discuss the special status of cardinal systems in the model (Section 3.3), followed by the possible significance of languages with hybrid systems of spatial orientation (Section 3.4).

3.1 The role of geography

The most striking result of the logistic regression analysis is the role of the geography factor, which exhibits significant correlations with almost all types of orientation systems. This pattern holds in both the combined and doubled approach, for both the full models (Tables 4 and 5 in the Appendix) and the reduced models with only economy and geography predictors (Tables 2 and 3). In fact, in the doubled method, inland geography exhibits statistically significant correlations with all five types of orientation systems, though the correlations are strongest with cardinal, elevation, and riverine systems; and the correlations are negative with coastal and land–sea* systems (Table 3). This confirms our intuition that orientation systems which are anchored in geographic features other than the land and the sea are more likely to be found in inland environments. On the other hand, the correlation between cardinal systems and inland geography is less expected, since we might expect cardinal systems to be more frequent among languages with diversified geography. This is likely to be an artifact of the doubled approach, give that under the combined approach, cardinal systems are negatively correlated with inland geography (log odds = –14.88). This suggests that in the doubled method the cardinal category is picking up numerous languages for which the secondary orientation axis is a cardinal axis.

Coastal geography is a much weaker predictor of orientation system type. In particular, coastal geography is not correlated with coastal orientation systems. In other words, while coastal geography may be a necessary condition for coastal orientation, it is clearly not sufficient.[6] This is likely because land–sea, which is also highly correlated with coastal geography, was the reference orientation type. This result also reflects the uneven distribution of orientation system types according to geography. As might be expected given the archipelagic nature of the Austronesian family, the vast majority (89 in combined model; 120 in doubled) of the languages in our sample were coded as having coastal geography, but only 23 (25 in doubled coding) of these exhibit coastal systems. In fact, all nine systems of spatial orientation (combined coding) are represented among languages with coastal geography. In contrast, among languages with inland geography only four of these systems are represented, and the overwhelming majority of these are riverine (see Figure 2).

Diversified geography is captured in the model as the intercept value, which shows strong negative correlations with elevation, land–sea*, land–sea + elevation and riverine + cardinal systems, and a slight positive correlation with cardinal systems (Table 2). This result is in line with our expectations that languages spread across both coastal and inland environments will find it difficult to rely on geographically anchored features for orientation. Hence, cardinal systems are more likely for such languages. Fully ten of the eighteen systems (56%) with diversified geography are cardinal. There is no correlation between diversified geography

6 All but three of the twenty-three languages with coastal systems were coded as having coastal geography. Those three—Alune, Buginese, and Makassarese—arguably could have been coded as coastal as well. However, we coded them as having diversified geography because all three have both inland and coastal populations.

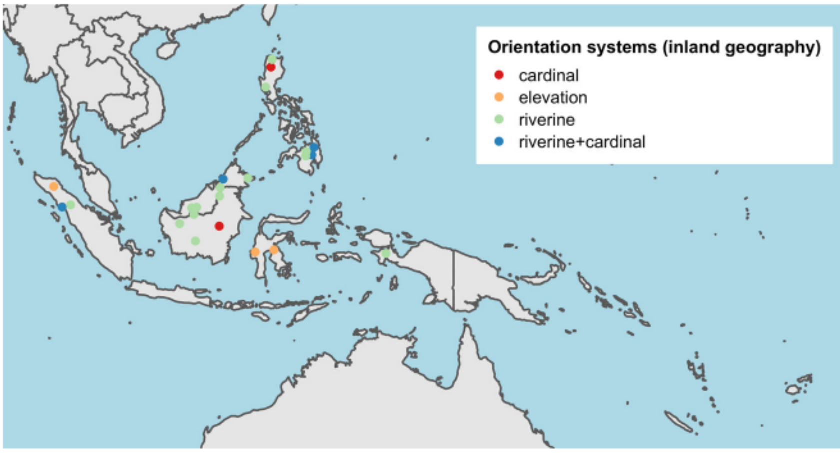

Figure 2: A restricted number of orientation systems are found among the twenty-three languages spoken in inland communities. Only four types are represented using our combined coding method.

and cardinal systems using the doubled method (Table 5), suggesting that several languages with village distribution and a secondary cardinal axis were picked up using this method.

Given that the simplified model without distribution and terrain factors performed slightly better than the model with those factors, we have so far focused our discussion of geographic influences on the geography predictor itself. However, given that both distribution and terrain are also geographic in nature, we briefly discuss the correlations found for those factors in the full model (Tables 4 and 5). The most surprising result is the almost complete lack of correlation with the terrain factor. Some (non-significant) correlations do indeed exist: nearly half (19 of 41) of languages spoken in non-mountainous environments have cardinal systems. However, all other types of systems are more commonly found in mountainous environments (see Figure 3).

The distribution factor presents a somewhat more interesting and less expected picture. Most notable is the strong correlation between distribution and both elevation and riverine + cardinal systems. All six languages with a riverine + cardinal system have a village distribution. Moreover, both island and village distribution are strongly correlated with elevation systems, whereas distributed distribution is negatively correlated. In part, this may reflect the areal concentration of elevation systems in Sulawesi—a large, linguistically diverse island where languages are usually localized to a small part of the island (cf. Holton and Pappas). As a result, nine of the fourteen elevation systems in the sample were coded as having a village distribution. This concentration is evident in the map in Figure 4, which shows the large number of elevation systems with village distribution in Sulawesi. However, elevation systems with village distribution are also found outside Sulawesi, as with Gayo, spoken in the interior of northern Sumatra. And not all Sulawesi languages with elevation systems have a village distribution, as with Muna, which has an island distribution and is spoken on the island of the same name in southeast Sulawesi.

3.2 The role of economy

The combined model (Table 2) provides evidence for a correlation between economy and orientation system type. Subsistence economy is negatively correlated with cardinal, land–sea + cardinal, and riverine + cardinal systems

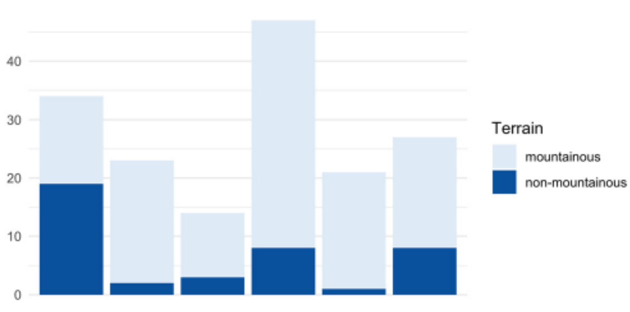

Figure 3: Frequency of mountainous and non-mountainous terrain according to orientation system type.

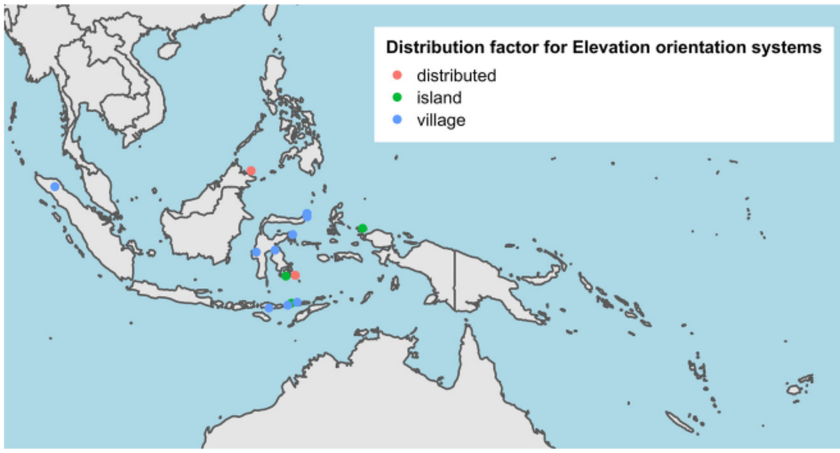

Figure 4: Locations of languages with elevation systems, indicating the value of the factor of distribution in each case.

but positively correlated with land–sea + elevation systems. The negative correlation of cardinal systems with subsistence economy is not surprising. We expect cardinal systems to be more likely among languages spoken in urban regions and in languages with large populations which will typically have a diversified economy. Diversified economy is found in twenty of the twenty-eight cardinal systems in the sample.

In the combined approach we see a strong correlation between subsistence economy and land–sea + elevation systems (log odds = 14.03), but this merely reflects the fact that all of the land–sea + elevation systems in the sample have subsistence economies. Likewise, a strong negative correlation with land–sea + cardinal systems (log odds = −24.77) reflects that two of the three languages in this category were non-subsistence.

In the doubled approach (Table 3) neither subsistence nor agriculture economy exhibits a significant correlation with any orientation type. In particular, there is no correlation between subsistence economy and non-cardinal orientation systems (i.e., no negative correlation between subsistence economy and cardinal systems). Note that several of the languages with subsistence economy have a secondary cardinal axis, and thus the coding for the doubled method contains relatively more instances of languages with subsistence economy and cardinal system. This may have the effect of weakening the correlation between subsistence economy and non-cardinal systems which was found using the combined method. However, subsistence economy is the overwhelmingly dominant economy type among land–sea systems, which represents the reference value for the model.

3.3 Cardinal systems

The intercept value in Table 2 indicates the likelihood of a language having a particular orientation system type as compared to a simple land–sea system, given the reference values for diversified economy and diversified geography. Only the cardinal system shows a significant positive correlation, suggesting a special status for such systems.

It is well known that cardinal systems in Austronesian languages tend to evolve from originally geomorphic or landmark-based spatial orientation systems (see Adelaar 1997; Gallego 2018). This is evidenced most clearly in the etymological sources for the cardinal axes, which commonly include topographic (e.g., land–sea) and environmental (e.g., path of the sun or wind) terms.[7] As language communities grow and become spread across large regions with diversified economies, geocentric axes tend to become conventionalized and reinterpreted as cardinal systems. Of the thirteen languages in our sample with diversified economy and diversified geography,

7 While we only treat cardinal systems based on topographic features as cardinal if they have been generalized to a fixed axis that is no longer rooted in topography, we treat all systems based on environmental factors—the path of the sun or the wind—as cardinal.

ten have simple cardinal systems, and two have hybrid elevation + cardinal systems. In other words, as populations become larger, more diversified, and less connected to the land, cardinal systems of orientation become more likely.

3.4 Languages with hybrid systems of orientation

Another notable result is the large number of significant correlations between the economy and geography predictors and the three hybrid orientation types (land–sea + cardinal, land–sea + elevation, and riverine + cardinal), shown in the three rightmost columns of Table 2. The number of hybrid systems is relatively small, representing only seventeen of the 131 languages in our sample data, but the correlations are nonetheless striking.

All languages make use of various strategies of spatial orientation, but when spatial orientation systems are anchored in (local) geography, the vast majority of those systems are anchored to just one type of feature. Thus, the choice to use more than one type of feature to anchor a system may well reflect sociotopographic factors. The seven languages with a hybrid land–sea + elevation system (Kedang, Lewotobi, Rongga, Solor, Waigeo, Bantik, and Toratán) are spoken in small language communities which tend to be reliant on subsistence farming and fishing. The six languages with a hybrid riverine + cardinal system (Jakarta Malay, Agusan Manobo, Ata Manobo, Central Subanen, Karo Batak, and Tumugon Murut) are largely localized to the larger islands of western Indonesia and the Philippines, where a village distribution is quite common and both riverine and cardinal systems are more prevalent. Finally, all three languages with a hybrid land–sea + cardinal system (Balinese, Sasak, and Tungag) are spoken minimally across an entire island and thus did not have inland geography. The small number of hybrid languages in our sample makes it difficult to make broader generalizations; yet results nonetheless reflect a striking similarity between the environments in which they are spoken.

Whether the presence or absence of a hybrid system is actually a significant typological characteristic remains to be seen. As shown in Figure 5, hybrid systems seem to cluster areally, which may explain the strong correlations found in the model. That said, as noted in Section 2, in order to reduce the number of coding points in our typology we greatly simplified the number of hybrid orientation types in the data, removing such hybrid systems as land–sea + coastal, elevation + cardinal, and elevation + riverine, and reclassifying these as simple systems. Including these systems in a hybrid versus simple typology results in doubling the number of hybrid systems (n = 35). The potential link between hybrid orientation systems and sociotopography merits further research.

4 Summary

Of the four sociotopographic features we examined in this study, two—economy and geography—showed statistically significant correlations with the type of spatial orientation system used in a language. In other words,

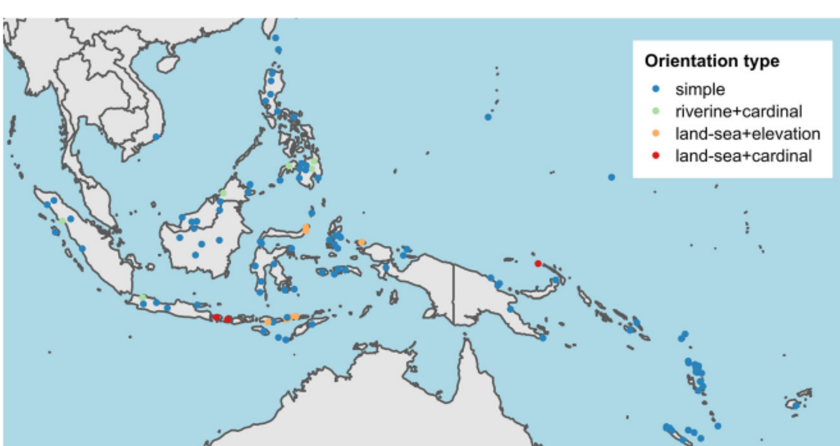

Figure 5: Distribution of languages with hybrid versus simple orientation systems.

geography alone is not sufficient to explain the distribution of spatial orientation systems: "even the simplest environmental influence is either supported or transformed by social forces" (Sapir 1912: 226). Despite the limitations of a small sample size, and the challenges of coding sociotopographic features, this preliminary work shows the potential for quantitative investigation of the Sociotopographic Model. In particular, these results support what might be called a "weak" form of the model. Sociotopographic features alone cannot completely predict the type of spatial orientation system in a language, but they may well be able to explain at least part of the observed variation in the anchoring of absolute orientation systems in Austronesian languages. To the extent that languages are constrained by grammar just as much as they are by culture and environment, spatial orientation systems are unlikely to be fully determined by language-external factors. Rather, a complex interplay of factors are at work (cf. Palmer et al. 2017). But statistical correlations still beg for explanation. As noted by Palmer (2015: 223), "if linguistic spatial systems correlate predictably with a pre-existing external world, then they must be constructed in response to that world in a process mediated by higher level cross-modal conceptual representations."

We emphasize that the models developed here are strictly exploratory in nature, and our choice to employ logistic regression was largely opportunistic. Given the small sample size and relatively large number of predictors, it would be useful to explore alternatives to regression methods, such as conditional inference trees and random forests.[8] In comparison to regression models, tree-based approaches perform better with categorical independent variables and are able to handle complex interaction effects (cf. Tagliamonte and Baayen 2012). Moreover, as noted in Section 2, in order to fit the model with a small amount of data we chose to simplify both the typology of spatial orientation systems and the number of categorical values for the independent variables. Given the typological nature of this study, these coding choices and simplifications necessarily rely heavily on subjective judgments and broad generalizations of data in secondary sources, many of which are not easy to interpret without firsthand knowledge of the language communities. Researchers with a deeper knowledge of the languages in question may have coded orientation systems or sociotopographic factors differently.[9] Further, it is worth pointing out that the sources consulted for this study vary greatly in both the depth and breadth of their treatment of spatial orientation. Some sources provide detailed descriptions of spatial orientation based on careful experimental work, while others merely provide a list of "directional" terms. In spite of these shortcomings, this study demonstrates the enormous potential for quantitative approaches to understanding the relationship between culture, environment, and spatial orientation.

Looking ahead, if Austronesian spatial orientation is indeed "a product of culture acting upon experience" (Blust 1997: 50), then a better understanding of that equation will require both detailed case studies of individual languages as well as broader quantitative studies of feature correspondence. In future work we hope to further expand our sample to better represent the vast diversity in spatial orientation systems among the Austronesian languages. In addition, we plan to continue to improve the sociotopographic variables to include continuous as well as polytomous factors (for example, terrain quantified to reflect the mean elevation change across a language area), while also attempting to make use of broadly available codings of economic and cultural factors, such as those represented in the Database of Places, Language, Culture, and Environment (D-PLACE, Kirby et al. 2016). In this way we hope to inspire additional quantitative work in sociotopography, both in Austronesian and beyond.

Acknowledgment: This paper benefited enormously from the feedback of two anonymous reviewers, as well as discussions with the volume editors. Any remaining errors or misrepresentations are solely the responsibility of the authors.
Research funding: This paper was supported in part by US National Science Foundation grants 1761223 and 2025315 (DOI: 10.13039/100000001).

8 Many thanks to the editors of this volume for this suggestion.
9 The coded database is included with the Supplementary Materials available at https://doi.org/10.5281/zenodo.4708029.

Appendix: Additional data

Tables 4 and 5 present the model results (log odds) using all four predictors, for the combined and doubled methods, respectively. Table 6 presents the values of Akaike's Information Criterion (AIC) for models run with all subsets of the four predictors. For both methods the lowest AIC, and hence best model fit, occurs with the economy and geography predictors.

Table 4: Log odds (combined method, all four predictors).

	Orientation type							
	Cardinal	Coastal	Elevation	Land–sea*	Riverine	Land–sea + cardinal	Land–sea + elevation	Riverine + cardinal
Island	−0.64	−2.49†	32.66†††	−0.9	−0.52	0.86	27.06	−4.63†††
	(0.58)	(0.05)	(0.00)	(0.46)	(0.78)	(0.62)	(0.83)	(0.00)
Village	−0.96	−1.04	32.74†††	−0.19	0.22	−39.63	26.54	32.58†††
	(0.46)	(0.35)	(0.00)	(0.87)	(0.92)	(0.00)	(0.83)	(0.00)
Agriculture	−0.71	−1.52	−2.18	−1.71	−1.12	−0.1	−18.83†††	−77.28†††
	(0.58)	(0.34)	(0.31)	(0.29)	(0.58)	(0.96)	(0.00)	(0.00)
Subsistence	−1.54	0.93	−2.66	0.42	−1.21	−39.19	12.14	−4.1††
	(0.22)	(0.48)	(0.14)	(0.75)	(0.52)		(0.99)	(0.03)
Coast	−1.93	−1.28	34.59†††	28.71†††	−3.08†	−1.97	32.05	10.58†††
	(0.13)	(0.37)	(0.00)	(0.00)	(0.08)	(0.3)	(0.9)	(0.00)
Inland	−32.54†††	−26.3	65.13†††	6.3	30.65†††	−2.07	2.95†††	42.23†††
	(0.00)		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Mountainous	−0.88	1.34	0.03	1.85	0.96	0.73	29.0	0.23
	(0.36)	(0.24)	(0.99)	(0.15)	(0.52)	(0.65)	(0.91)	(0.91)
Intercept	3.48††	0.99	−66.83†††	−29.54†††	0.8	0.31	−100.2	−41.33†††
	(0.01)	(0.53)	(0.00)	(0.00)	(0.65)	(0.88)	(0.68)	(0.00)

p-values in parentheses. AIC: 363.480, McFadden's R^2: 0.556. †$p < 0.1$. ††$p < 0.05$. †††$p < 0.01$.

Table 5: Log odds (doubled method, all four predictors).

	Orientation type				
	Cardinal	Coastal	Elevation	Land–sea*	Riverine
Island	−0.34	−1.78†	16.47†††	−0.12	−0.00
	(0.7)	(0.08)	(0.00)	(0.9)	(1.0)
Village	0.04	−0.54	15.65†††	0.1	2.18
	(0.97)	(0.48)	(0.00)	(0.91)	(0.15)
Agriculture	1.0	0.96	0.24	0.93	0.53
	(0.29)	(0.49)	(0.89)	(0.52)	(0.68)
Subsistence	−1.8†	1.45	−0.76	0.95	−2.15†
	(0.08)	(0.24)	(0.59)	(0.44)	(0.09)
Coast	1.19	0.92	−16.42†††	−15.41†††	1.62
	(0.18)	(0.43)	(0.00)	(0.00)	(0.23)
Inland	−32.54	−3.53†††	16.28†††	−3.77†††	18.72†††
	(0.99)	(0.00)	(0.00)	(0.00)	(0.00)
Mountainous	0.97	−0.85	−0.71	−1.27	0.47
	(0.2)	(0.39)	(0.6)	(0.29)	(0.67)
Intercept	−0.64	−1.07	−16.54†††	−1.11	−2.71†
	(0.47)	(0.36)	(0.00)	(0.35)	(0.09)

p-values in parentheses. AIC: 373.417, McFadden's R^2: 0.498. †$p < 0.1$. ††$p < 0.05$. †††$p < 0.01$.

Table 6: Akaike's Information Criterion for subsets of the four predictors.

Subset of predictors	Combined	Doubled
All predictors	363.48	373.42
Predictors excluding distribution	351.50	370.56
Predictors excluding economy	468.97	486.37
Predictors excluding geography	367.73	386.74
Predictors excluding terrain	358.38	369.67
Distribution and economy	367.23	386.90
Distribution and geography	478.25	497.76
Distribution and terrain	516.16	551.83
Economy and geography	347.93	368.84
Economy and terrain	374.52	397.29
Geography and terrain	454.09	479.38
Distribution only	528.05	567.83
Economy only	376.07	399.25
Geography only	473.41	505.37
Terrain only	525.41	564.46

References

Adelaar, K. Alexander. 1997. An exploration of directional systems in West Indonesia and Madagascar. In Gunter Senft (ed.), *Referring to space: Studies in Austronesian and Papuan languages* (Oxford Studies in Anthropological Linguistics 11), 53–82. Oxford: Oxford University Press.

Blust, Robert. 1997. Semantic change and the conceptualization of spatial relationships in Austronesian languages. In Gunter Senft (ed.), *Referring to space: Studies in Austronesian and Papuan languages* (Oxford Studies in Anthropological Linguistics 11), 39–52. Oxford: Oxford University Press.

Blust, Robert. 2013. *The Austronesian languages*. Canberra: Pacific Linguistics.

Bohnemeyer, Jürgen, Katharine T. Donelson, Randi E. Moore, Elena Benedicto, Alyson Eggleston, Carolyn K. O'Meara, Gabriela Pérez Báez, Alejandra Capistrán Garza, Néstor Hernández Green, María de Jesús Selene Hernández Gómez, Samuel Herrera Castro, Enrique Palancar, Gilles Polian & Rodrigo Romero Méndez. 2015. The contact diffusion of linguistic practices: Reference frames in Mesoamerica. *Language Dynamics and Change* 5(2). 169–201.

Bowden, John. 1992. *Behind the preposition: Grammaticalisation of locatives in Oceanic languages* (B-107). Canberra: Pacific Linguistics.

Bubandt, Nils. 1997. Speaking of places: Spatial poesis and localized identity in Buli. In James J. Fox (ed.), *The poetic power of place*, 131–162. Canberra: Research School of Pacific and Asian Studies.

Gallego, Maria Kristina S. 2018. Directional systems in Philippine languages. *Oceanic Linguistics* 57(1). 63–100.

Graul, Christian. 2016. leafletR: Interactive web-maps based on the Leaflet JavaScript Library. R package version 0.4-0. Available at: http://cran.r-project.org/package=leafletR.

Hammarström, Harald, Robert Forkel & Martin Haspelmath & Sebastian Bank. 2020. Glottolog 4.2.1. Max Planck Institute for the Science of Human History. Available at: https://glottolog.org.

Heegård, Jan & Henrik Liljegren. 2018. Geomorphic coding in Palula and Kalasha. *Acta Linguistica Hafniensia* 50(2). 129–160.

Holton, Gary. 2017. A unified system of spatial orientation in the Austronesian and non-Austronesian languages of Halmahera. *NUSA* 62. 157–189.

Holton, Gary & Leah Pappas. to appear. Spatial orientation in the Malayo-Polynesian languages outside Oceanic. In K. Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the Malayo-Polynesian languages of Southeast Asia*. Oxford: Oxford University Press.

Kirby, Kathryn R., Russell D. Gray, Simon J. Greenhill, Fiona M. Jordan, Stephanie Gomes-Ng, Hans-Jörg Bibiko, Damián E. Blasi, Carlos A. Botero, Claire Bowern, Carol R. Ember, Dan Leehr, Bobbi S. Low, Joe McCarter, William Divale & Michael C. Gavin. 2016. D-PLACE: A global database of cultural, linguistic and environmental diversity. *PLOS ONE* 11(7). e0158391.

Levinson, Stephen C. 2003. *Space in language and cognition*. Cambridge: Cambridge University Press.

Palmer, Bill. 2002. Absolute spatial reference and the grammaticalisation of perceptually salient phenomena. In Giovanni Bennardo (ed.), *Representing space in Oceania: Culture in language and mind*, 107–157. Canberra: Australian National University.

Palmer, Bill. 2015. Topography in language: Absolute frame of reference and the Topographic Correspondence Hypothesis. In Rik De Busser & Randy J. LaPolla (eds.), *Language structure and environment: Social, cultural, and natural factors*, 179–226. Amsterdam: John Benjamins.

Palmer, Bill, Jonathon Lum, Jonathan Schlossberg & Alice Gaby. 2017. How does the environment shape spatial language? Evidence for sociotopography. *Linguistic Typology* 21(3). 457–491.

Sapir, Edward. 1912. Language and environment. *American Anthropologist* 14(2). 226–242.

Schlossberg, Jonathan. 2018. *Atolls, islands, and endless suburbia: Spatial reference in Marshallese*. Newcastle, Australia: The University of Newcastle PhD thesis.

Senft, Gunter (ed.). 1997. *Referring to space: Studies in Austronesian and Papuan languages* (Oxford Studies in Anthropological Linguistics 11). Oxford: Oxford University Press.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178.

Venables, William N. & Brian D. Ripley. 2002. *Modern applied statistics with S*, 4th edn. New York: Springer.

Wassmann, Jürg & Pierre R. Dasen. 1998. Balinese spatial orientation: Some empirical evidence of moderate linguistic relativity. *The Journal of the Royal Anthropological Institute* 4. 689–711.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/lingvan-2020-0044).