PNAS
www.pnas.org

Supplementary Information for

The forms and meanings of grammatical markers support efficient communication

Francis Mollica, Geoff Bacon, Noga Zaslavsky, Yang Xu, Terry Regier and Charles Kemp

Francis Mollica.
E-mail: fmollica@ed.ac.uk

This PDF file includes:

Supplementary text
Figs. S1 to S12 (not allowed for Brief Reports)
Tables S1 to S9 (not allowed for Brief Reports)
SI References

Supporting Information Text

1. Representing grammatical systems

As described in the main text, for each grammatical feature we define an underlying semantic dimension that specifies possible states of the world. For each attested system, we define the language's encoder $q_m(m|s)$ using a maximum entropy assumption over the meanings m that can be used to convey a speaker distribution s. For example, in the Sanskrit number system two objects can only be communicated using the dual so $q(\text{DUAL}|s_2) = 1$. In Slovene, however, the dual is optional and so s_2 can be expressed using either the dual or the plural. In this case, the maximum entropy assumption dictates that $q(\text{DUAL}|s_2) = q(\text{PL}|s_2) = 0.5$. Section B includes an analysis suggesting that our conclusions are robust to variations in the maximum entropy assumption.

When enumerating hypothetical systems (grey dots in Figure 2) we include all systems that partition the underlying semantic dimension. Systems with optional markers could be considered in principle, but doing so would make the analysis intractable. In general, allowing the same state to be labelled with two or more markers decreases both complexity and communicative precision. As a more stringent test of our theory, section C compares attested systems against permutations of attested systems, and these permutations represent a subset of all possible optional partitions. Further, as attested languages often partition the semantic dimension into connected regions, section D repeats our comparisons of attested systems against all possible partitions and against permutations of attested systems but imposes an additional semantic connectivity constraint.

As described in the main text, temporal adverbs used to estimate the prior $P(s)$ for tense are listed in Table S1.

Deg. of remoteness	**Temporal adverbs**
immediate	*today*
near past	*yesterday*
near future	*tomorrow*
remote past	*last week/month/year/decade/century*
remote future	*next week/month/year/decade/century*

Table S1. Temporal adverbs used to estimate usage probabilities for locations along the timeline. Because the immediate past and immediate future are both expressed through *today* in English, we assigned half of *today*'s frequency to each of the two temporal locations.

We coded the forms of grammatical markers f in order to explore the relationship between empirical and optimal codelengths. Grammatical features are often realized within a paradigm, which can result in multiple forms for the same grammatical feature value (see Table S2). For our analyses, the observed length of a feature value is defined as an average over the lengths in characters for all forms of that feature value. The observed lengths in Figure 4 are normalized for each language by dividing by the maximum observed length for that language.

We gathered forms for grammatical features using monographs on each grammatical feature as a starting point (number: (1), tense: (2), evidentiality (3)). If the monograph contained forms for a system, we used those forms. Additionally, we sought out source grammars, where possible, for each language to extract forms from a primary source. To allow a strong test of our theory we focused on languages with a relatively large number of forms. Tense and evidentiality are predominantly realized as grammatical morphemes that attach to the verb. The realization of grammatical number is more varied across languages. Number can be marked on nouns, pronouns, verbs, adjectives, determiners, case markers and more. Further, number marking can be lexicalized or realized via additive morphology (affixes), reduplication, or suppletion (for a survey of grammatical pluralization strategies see (4)). We therefore compiled a list of all possible ways that number is marked in our sample of languages (available on OSF). Based on this survey, we chose to focus on nominal and pronominal marking in order to cover as many languages as possible from our sample (33 out of 37). In doing so our approach differs from previous studies that have focused either on the marking of specific feature values (e.g., plurality (4)) or on the notion of "markedness" (5, 6). Nonetheless, we arrive at similar conclusions.

Future		First	Second	Third
	SG	*habl**aré***	*habl**arás***	*habl**ará***
	PL	*habl**aremos***	*habl**aréis***	*habl**arán***
Present	SG	*habl**o***	*habl**as***	*habl**a***
	PL	*habl**amos***	*habl**áis***	*habl**an***
Past Perfect	SG	*habl**é***	*habl**aste***	*habl**ó***
	PL	*habl**amos***	*habl**astais***	*habl**aron***
Past Imperfect	SG	*habl**aba***	*habl**abas***	*habl**aba***
	PL	*habl**ábamos***	*habl**abais***	*habl**aban***

Table S2. Example paradigm for for tense forms in Spanish.

2. Alternate Measures of Complexity

Our metric of complexity is based on an information-theoretic analysis of communication, but linguists have proposed many alternative metrics of morphological complexity (e.g., 7, 8). Nichols (9) discusses two general ways of measuring complexity:

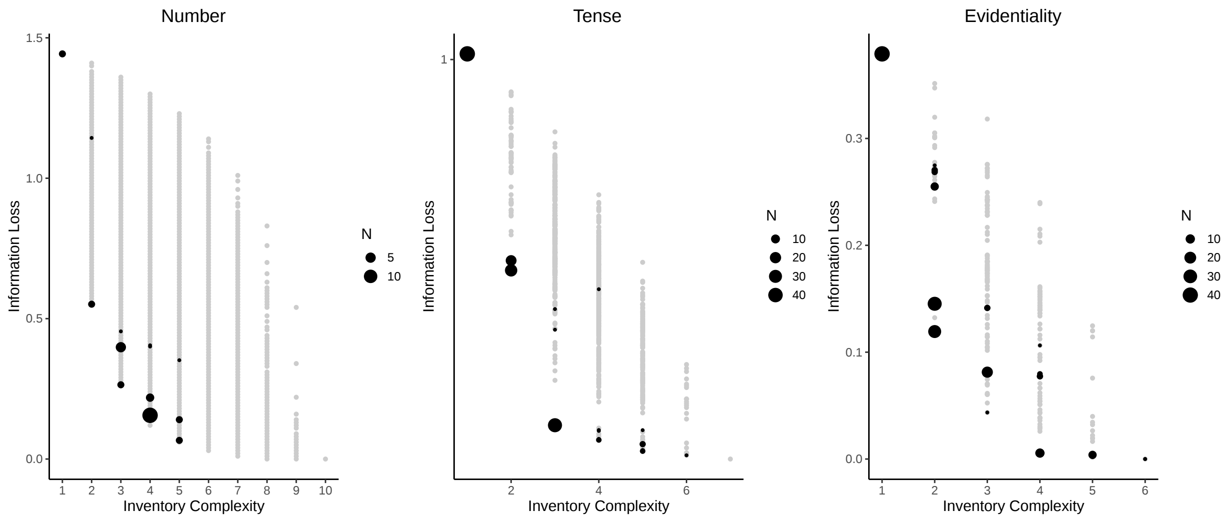

Fig. S1. Trade-off between information loss and inventory complexity.

inventory complexity, or the number of unique morphemes in a system, and descriptive complexity (Kolmogorov Complexity), or the minimal amount of information required to describe a system. Figure S1 shows the tradeoff between information loss and complexity when complexity is defined as inventory complexity, and the same approach was previously applied to tense marking in (10). The results still provide some evidence that attested systems of a given complexity level tend to minimize information loss, which is not surprising as measures of morphological complexity tend to be correlated (11). The results, however, reveal that replacing our information-theoretic complexity measure (Equation 1) with inventory complexity means that departures from the Pareto frontier become more common and more substantial, especially for systems with low inventory complexity. The results therefore suggest that our model accounts better for attested systems than a similar approach that relies on inventory complexity.

Some previous work that explores tradeoffs between information loss and complexity has relied on descriptive complexity (12–14), and a similar approach could be applied here. Previous work in this tradition typically commits to a domain-specific representation language, and complexity is then defined as the length of the shortest representation in that language. Extending this approach to our setting would probably require three separate representation languages for number, tense and evidentiality. Formulating these representation languages seems relatively challenging, and formulating the domain-specific components of our approach (the speaker distributions s) seems straightforward by comparison.

An alternative approach is to invoke a universal representation language with the property that the length of an object's representation is inversely proportional to its probability. Approaching the problem in this way allows theoretical connections to be established between descriptive complexity and information-theoretic measures of complexity (15). From this perspective, our complexity measure defined in Equation 1 can be viewed as a kind of descriptive complexity measure.

3. Quantitative Analyses

To further explore the results in Figures 2, 3 and 4, we ran a series of quantitative evaluations.

A. Evaluation of Near-Optimality. Figure 2 suggests that attested systems achieve near-optimal tradeoffs between the dimensions of information loss and complexity. A tradeoff implies that both dimensions matter, and the signature of this tradeoff is that attested systems tend to lie near the Pareto frontier. The penultimate column in Table 1 shows the Euclidean distance between a system and our 2-dimensional Pareto frontier, but here we also consider the generalized Normalized Information distance (gNID), an alternative measure introduced in (16). Considering gNID allows us to directly quantify the similarity between attested systems and their closest optimal counterparts. For completeness, we briefly present the formal definition of gNID (see the SI of (16) for more detail). Let m_1 and m_2 be markers drawn from encoders q_1 and q_2 respectively, where

$$q(m_1, m_2) = \sum_s p(s) q_1(m_1|s) q_2(m_2|s). \quad [1]$$

The distance between q_1 and q_2 is low if the mutual information $I(M_1; M_2)$ is high, and the gNID is calculated using

$$\text{gNID}(q_1, q_2) = 1 - \frac{I(M_1; M_2)}{\max\{I(M_1; M_1'), I(M_2; M_2')\}}. \quad [2]$$

If attested system q resembles a system q^* on the Pareto frontier, then $\text{gNID}(q, q^*)$ will be low. We calculate gNID for each attested system q with respect to the optimal system q^* that minimizes gNID.

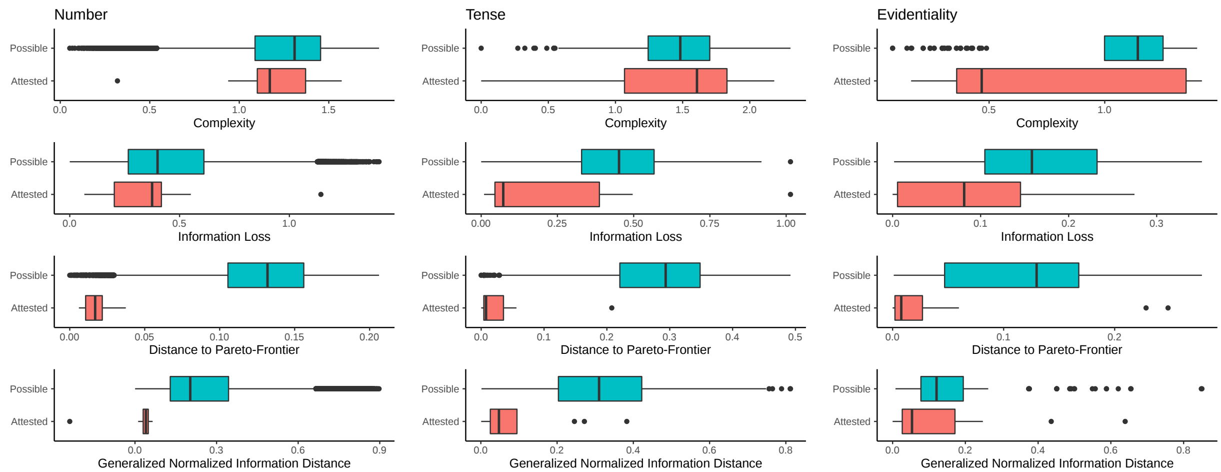

Fig. S2. The distributions of different evaluation metrics for unattested and attested grammatical systems.

Figure S2 compares attested and unattested inventories with respect to four evaluation metrics: complexity, information loss (KL Divergence), distance to the Pareto frontier, and generalized normalized information distance (gNID). Comparing these evaluation metrics allows us to see whether attested inventories are near-optimal with respect to one dimension alone, or whether attested inventories navigate a tradeoff between information loss and complexity.

	Number	Tense	Evidentiality
Complexity	-7632 (15.3)	-61.5 (0.040)	-12.6 (0.128)
Information Loss	-7986 (10.1)	-48.5 (0.313)	-13.9 (0.070)
Distance to Pareto Frontier	**-547 (7.02)**	**-16.9 (0.500)**	**-12.0 (0.233)**
gNID	-3162 (42.6)	-39.5 (0.453)	-14.5 (0.064)

Table S3. Model comparison using log likelihood values averaged across CV runs. Higher values are considered better fit. Standard errors in parentheses.

To quantitatively test for near optimality, we fit four logistic regression models using each metric to predict whether or not a system is attested. To adjust for the imbalance in number of attested and possible systems, we over-sampled the attested systems (17) and compared models using the average log likelihood from 10 runs of 5-fold cross-validation, repeated 10 times. All models were implemented in R (18) using the `tidymodels` (19), `themis` (20) and `lme4` (21) packages. As can be seen in Table S3, Euclidean distance to the Pareto frontier better predicts the diversity of attested grammatical systems than either dimension of the trade-off alone.

		Number	Tense
Favored large	Complexity	-7591 (15.3)	-59.7 (0.035)
	Information Loss	-8021 (6.61)	-50.7 (0.31)
	Distance to Pareto Frontier	**-339 (5.44)**	**-17.1 (0.55)**
Disfavored large	Complexity	-7783 (13.2)	-59.6 (0.04)
	Information Loss	-8027 (6.67)	-50.9 (0.30)
	Distance to Pareto Frontier	**-674 (9.43)**	**-17 (0.55)**

Table S4. Model comparison with weighted optional encoding that either favors ($\approx 75\%$) or disfavors ($\approx 25\%$) the larger partition. Values are log likelihoods averaged across CV runs. Higher values are considered better fit. Standard errors in parentheses.

B. Relaxing the maximum entropy assumption for optional distinctions. To assess the sensitivity of our conclusions to our max entropy assumption, we replicated our near optimality analysis assigning weights that either favor the larger partition ($\approx 75\%$) or disfavor the larger partition ($\approx 25\%$). Regardless of the weighting, Table S4 shows that Euclidean distance to the Pareto frontier better predicts the diversity of attested grammatical systems than either dimension of the trade-off alone.

C. Permutation Analysis for Near-Optimality. As mentioned above, our hypothetical systems included all possible partitions of the semantic space. Some of these systems may be linguistically implausible, including the number system with ten different meanings associated with each of the ten world states. We therefore conducted a second analysis using a more restricted set of comparison systems that includes only permutations of attested systems (22). Each permuted system is created by starting with an attested system and then permuting the underlying semantic space. Figure S3 is similar to Figure 2 but

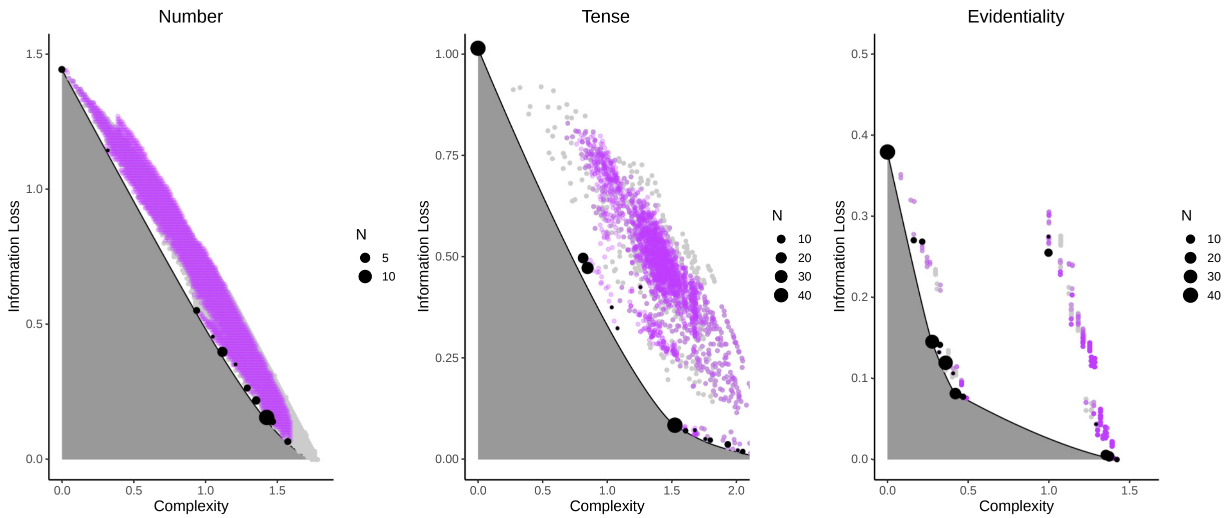

Fig. S3. Permutation analysis of the meaning of grammatical markers. Top panels: (a)-(c) Trade-offs between information loss and complexity for number, tense and evidentiality. Attested inventories (black points), permuted systems (purple points) and unattested systems (grey points) are plotted in the space of all possible grammatical systems. Systems that achieve optimal trade-offs lie along the Pareto frontier (solid line), and the shaded region below the line shows trade-offs that are impossible to achieve.

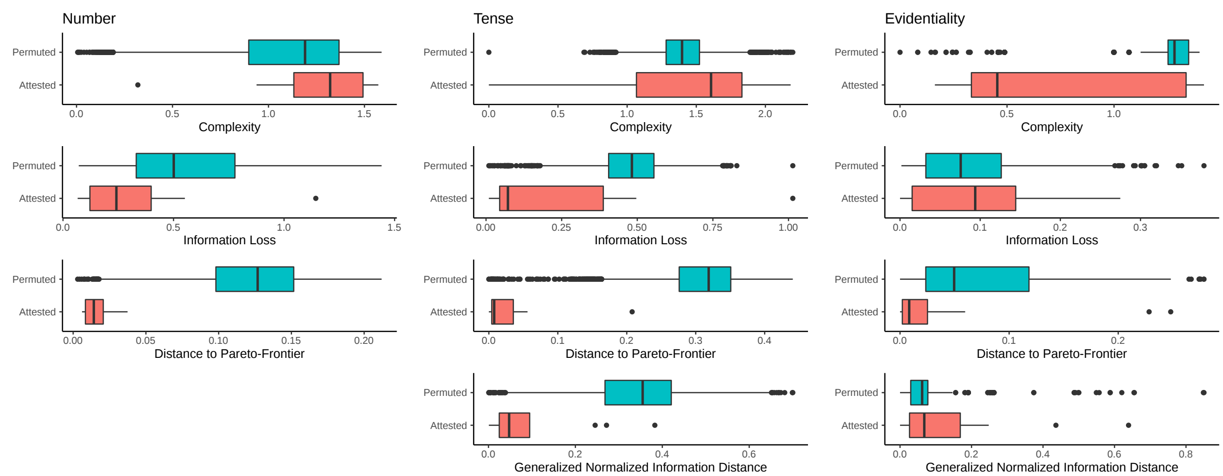

Fig. S4. The distributions of different evaluation metrics for permuted and attested grammatical systems.

shows all possible permutations in purple, and Figure S4 compares attested and permuted inventories with respect to our evaluation metrics. We leave out the gNID predictor for number, as calculating gNID for all possible permutations would be computationally expensive.

Again, to quantitatively test for near optimality, we fit four logistic regression models using each metric to predict whether or not a system is attested. To adjust for the imbalance in number of attested and permuted systems, we over-sampled the attested systems (17) and compared models using the average log likelihood from 10 runs of 5-fold cross-validation, repeated 10 times. All models were implemented in R (18) using the `tidymodels` (19), `themis` (20) and `lme4` (21) packages. As can be seen in Table S5, Euclidean distance to the Pareto frontier better predicts the diversity of attested grammatical systems than either dimension of the trade-off alone for number and tense. For evidentiality, complexity alone emerges as the best predictor.

D. Adding a semantic connectivity constraint. We can call a system *connected* if it partitions a semantic space into a set of connected regions. The attested languages in our tense and number samples are all connected, but the analyses in Tables S3 and S5 both compare attested systems against alternatives that mostly do not satisfy this constraint. To enable stronger tests of our theory we therefore repeated these analyses but filtered the comparison set (either all possible partitions or permutations of attested systems) to retain only connected systems. As seen in Table S6, for number and tense Euclidean distance to the Pareto-frontier was a better predictor of attested systems than either dimension alone. For evidentiality, information loss was the best predictor for distinguishing between possible and permuted systems. This result therefore suggests that evidentiality systems may primarily be driven by minimizing information loss under the constraint that feature values correspond to connected regions of the semantic space. The permutation analysis without the connectivity constraint indicated

	Number	Tense	Evidentiality
Complexity	-19484 (27.1)	-159 (0.195)	**-39.8 (0.331)**
Information Loss	-17567 (58.8)	-114 (0.786)	-50.8 (0.125)
Distance to Pareto Frontier	**-1124 (11.5)**	**-36.9 (0.840)**	-49.3 (0.184)
gNID	NA	-81 (0.912)	-48.5 (0.156)

Table S5. Permutation model comparison using log likelihood values averaged across CV runs. Higher values are considered better fit. Standard errors in parentheses.

		Number	Tense	Evidentiality
Possible Systems	Complexity	-38.6 (0.07)	-5.81 (0.05)	-2.87 (0.13)
	Information Loss	-37.2 (0.07)	-5.01 (0.08)	**-2.58 (0.05)**
	Distance to Pareto Frontier	**-22.9 (0.47)**	**-4.30 (0.23)**	-3.18 (0.13)
	gNID	NA	-4.47 (0.19)	-2.88 (0.16)
Permuted Systems	Complexity	-42.6 (0.32)	-8.75 (0.05)	-9.92 (0.16)
	Information Loss	-45.0 (0.33)	-10.2 (0.07)	**-9.3 (0.07)**
	Distance to Pareto Frontier	**-30.0 (0.44)**	**-7.85 (0.22)**	-11.1 (0.05)
	gNID	NA	-7.92 (0.11)	-11.8 (0.06)

Table S6. Model comparison with a semantic connectivity constraint. Values are log likelihoods averaged across CV runs. Higher values are considered better fit. Standard errors in parentheses.

that complexity was the best predictor of evidentiality systems (Table S5), but adding the connectivity constraint reduces variation in complexity and therefore yields a different conclusion.

E. Evaluation of Tradeoff between Zero Marking and Information Loss. To test whether information loss influences the pattern of zero-marking in tense systems, we ran a logistic mixed effect regression model predicting zero-marking according to (2) (present/absent) as function of information loss with a random intercept for language family. The model was implemented in R (18) using the `lme4` package (21). For all `lme4` model comparisons, p-values are calculated from a likelihood ratio test between models with and without the predictor of interest. Languages with maximal simplicity (i.e. with only one feature value) were removed for this analysis. Languages are more likely to use zero-marking at high information loss than at low information loss ($\beta = 10, \mathrm{SE} = 2.3, p < 0.05$). Because our sample was not chosen to reflect typological frequency of all attested tense systems, this finding should be treated as preliminary.

Figure S5 is analogous to Figure 3 but is based on the subset of our evidentiality dataset for which we have forms. This subset includes 15 languages with zero-marking and 16 languages without zero marking. A language was classified as zero-marked if there was at least one uninflected grammatical feature value. As for tense, we find that languages are more likely to use zero-marking at high information loss than at low information loss ($\beta = 15.6, \mathrm{SE} = 4.6, p < 0.05$) Unlike tense, however, we specifically selected the subset of evidentiality languages to favor languages with explicit forms. The results of the regression analysis for evidentiality should therefore be interpreted with caution.

Because we included both nominal and pronominal forms for grammatical number, our sample of number systems did not include zero marking. Previous work, however, suggests that the singular is sometimes zero marked but that zero marking of other feature values is extremely rare (23). A standard explanation for this finding is that zero marking tends to be applied to the most frequent feature value (here the singular) (24, 25), and the same explanation follows from our theoretical framework.

F. Evaluation of Correlation between Forms and Optimal Codelengths. Figure 4 compares optimal and observed form lengths across all languages for which we compiled forms, and results for individual languages are shown in Figures S6 - S8.

To test for a relationship between observed and optimal codelengths, we conducted separate linear mixed effect regression models for each grammatical feature predicting observed length for each feature value as a function of optimal codelength with random intercepts for Language and Language Family. The models were implemented in R (18) using the `lme4` package (21). We find a significant linear relationship between optimal codelengths and observed form lengths for number($\beta = 0.29, \mathrm{SE} = 0.05, p < 0.05$), tense ($\beta = 0.55, \mathrm{SE} = 0.08, p < 0.05$) and evidentiality ($\beta = 0.47, \mathrm{SE} = 0.06, p < 0.05$).

4. Individual Language Analyses

As we move from top left (minimal complexity) to bottom right (maximal complexity), along the IB Pareto frontier, the effective number of feature values in near-optimal systems gradually increases (26). Using the methods developed in (26), we can identify and analyze these structural changes as complexity is increased. This type of analysis was proposed and subsequently tested as a predictive model for the evolutionary trajectory of color naming systems (16, 26). For example, the model recapitulates a set of well-known typological claims (27), such as if a language has a category for green then it must have a category for red. This type of analysis has also been used to account for typological claims about an implicational evolutionary hierarchy of animal taxonomies (22), suggesting that this theoretically-driven approach may apply more broadly to the evolution of the lexicon (26). In our context, similar typological statements have been made for the grammatical features in our analysis. For example, if a language makes a dual distinction in number it must make a singular and plural distinction.

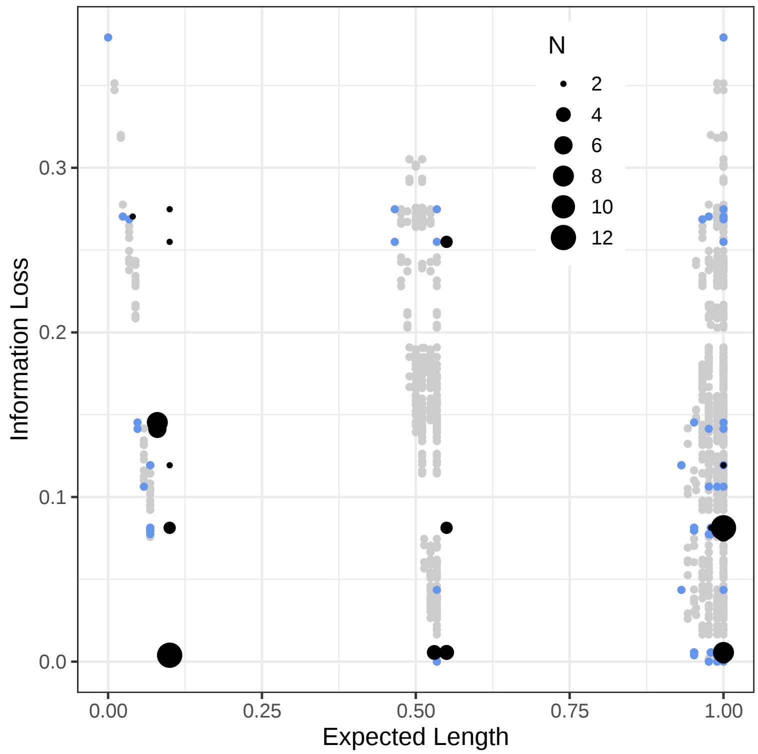

Fig. S5. Tradeoff between expected length and information loss for evidentiality. The size of the black dots reflect number of attested systems. The blue dots show all ways to zero-mark at most one feature value in an attested system. The grey dots show all ways to apply zero-marking to unattested systems. The column of attested systems with expected length equal to 1 includes systems that do not use zero marking.

Traversing the IB Pareto frontier in our case will generate a theoretical trajectory for grammatical features. However, the optimal systems along this frontier are generally stochastic (28), while systems for attested grammatical features are often deterministic. We will therefore also examine the deterministic systems for each number of distinctions that are the closest to optimal systems along the frontier. We use gNID as the distance measure to identify the optimal system closest to each deterministic system.

In this section, we visualize several optimal systems, together with their deterministic approximations, at discrete points along the IB frontier, as the number of terms increases. We then qualitatively compare these theoretical systems with individual languages in our sample. In each case we omit discussion of the least complex system, which has a single feature value, and the most complex system, which has unique feature values for every state of the world. We connect these trajectories with typological claims about grammatical features, but make no strong claims about diachronic change given that our results are based on synchronic data. Although we organize our discussion around a sequence of systems that increase in complexity, we do not claim that evolutionary forces always act in this direction, and do not commit to any single mechanism that may serve to increase or decrease complexity. For example, there may be directional changes in the mode of expression (syntactic forms becoming morphological) or there may be changes in our assumptions about communicative need or speaker distributions over time and language contact. Closer modelling of the pathways of grammaticalization (e.g., as described by 29) is needed in future research.

For ease of exposition, we highlight clear categorical predictions of the model (e.g. the past tense should be subdivided before the future). These predictions, however, are subject to uncertainty about the assumptions made when applying the model. For example, the predicted asymmetry between past and future is induced by an asymmetry in usage probabilities for past (0.274) compared to future (0.251), and would reverse if these probabilities were exchanged. Several corpus analyses suggest that the past is mentioned more frequently than the future (24, 30), but other aspects of our approach have a less firm empirical grounding. In particular, as suggested earlier the hierarchy assumed for evidentiality and the prior over this hierarchy should both be viewed as provisional. This section lays out the fine-grained predictions that result from our best attempt to ground the theoretical framework in available empirical data, but some of these predictions may need to be revised as better characterizations of speaker distributions and usage frequencies become available.

A. Number. Previous work on grammatical number has proposed a Number Hierarchy*:

$$\text{SG} > \text{PL} > \text{DU} > \text{TR}, \qquad [3]$$

*We adopt our exposition and notation from Corbett (1).

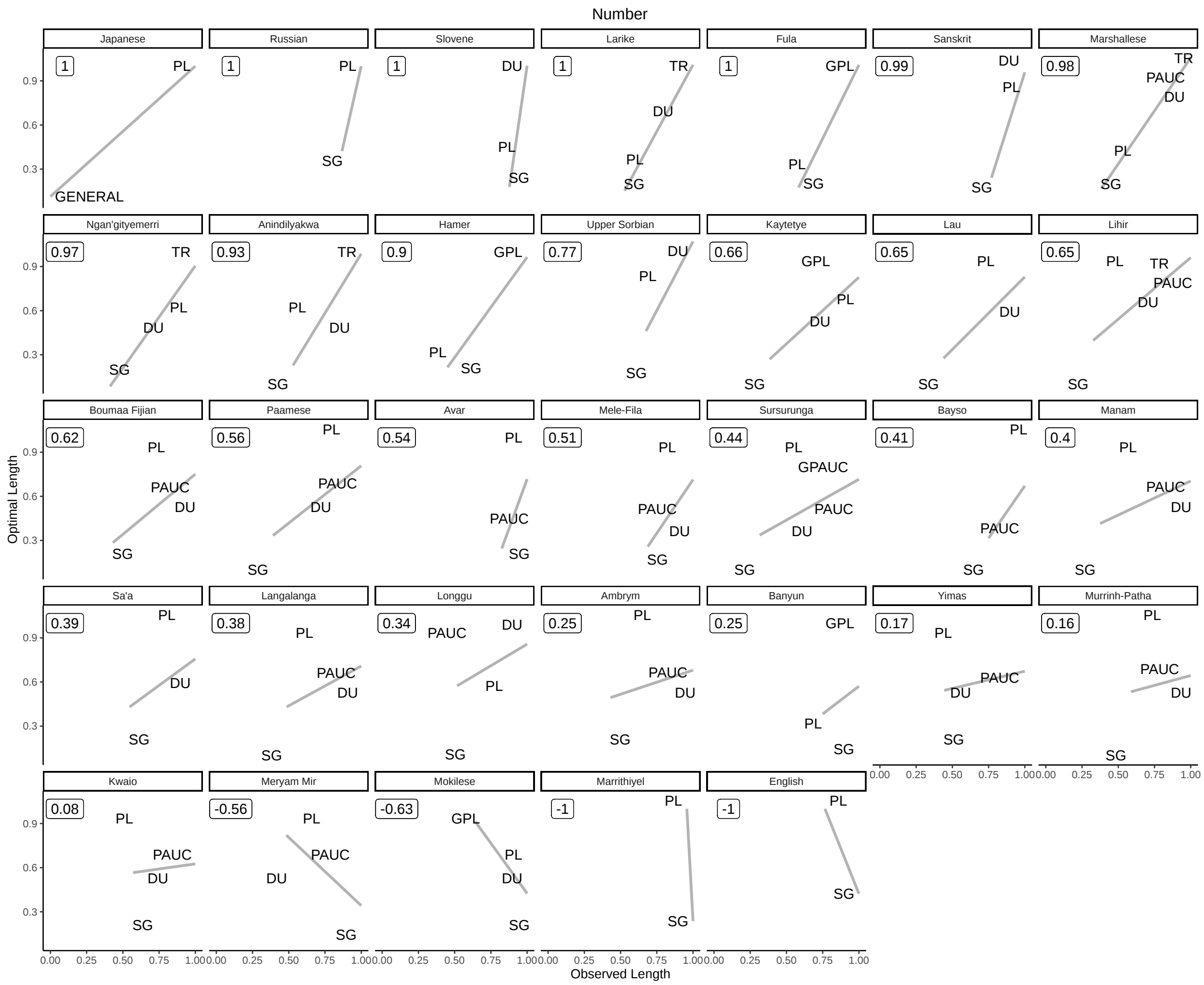

Fig. S6. Optimal lengths against observed lengths for individual languages in our number data set. Correlations are shown at the top left of each panel.

most likely hailing from Greenberg (23, pg. 94)'s universal 34:

"No language has a trial number unless it has a dual. No language has a dual unless it has a plural."

As the original universal does not mention the paucal, Croft (31) and Foley (32) have proposed to extend the hierarchy to:

$$\text{SG} > \text{PL} > \text{DU} > \text{TR}/\text{PAUC}. \quad [4]$$

Corbett (1) argues that this would still be insufficient to explain the world's languages. Under (4), the presence of a paucal would imply the presence of a dual. This does not occur in Bayso, which distinguishes SG, PAUC, PL. Corbett (1) also notes that number hierarchies make incorrect predictions when a feature value is optional. For example, in Serbian, the dual is optional so a speaker has the choice to use DU or PL morphology when talking about their *two* feet. For comparison, Sanskrit has the same "feature values" as Serbian; however, the speaker can only use the dual when talking about their *two* feet. Ostensibly under (4), an optional value should result in a choice between the value and its immediate predecessor in the hierarchy. While this works for Slovenian, it does not work for systems like Ngan'gityemerri, which distinguishes between SG, DU, TR and PL and has an optional trial. While (4) predicts a choice between TR and DU, the immediate predecessor, speakers of Ngan'gityemerri actually choose between TR and PL.

To account for these data, Corbett (1) proposes a binary branching structure, where different languages "choose" how much they carve out of the plural. If a language carves out a determinate amount–i.e., exact value, it must follow the traditional number hierarchy (3). At any step, a language could choose to carve out an indeterminate amount (e.g., a paucal). Under this account, optional distinctions can be made by collapsing the lowest distinctions back into the plural. This collapse can occur multiple times in a language. For example, Larike has SG, DU, TR and PL values but both the dual and the trial can be used optionally with the plural. Using this typology, all of the data can be accommodated; however, the exact trajectory by which a language would increase its set of feature values is less precisely stated.

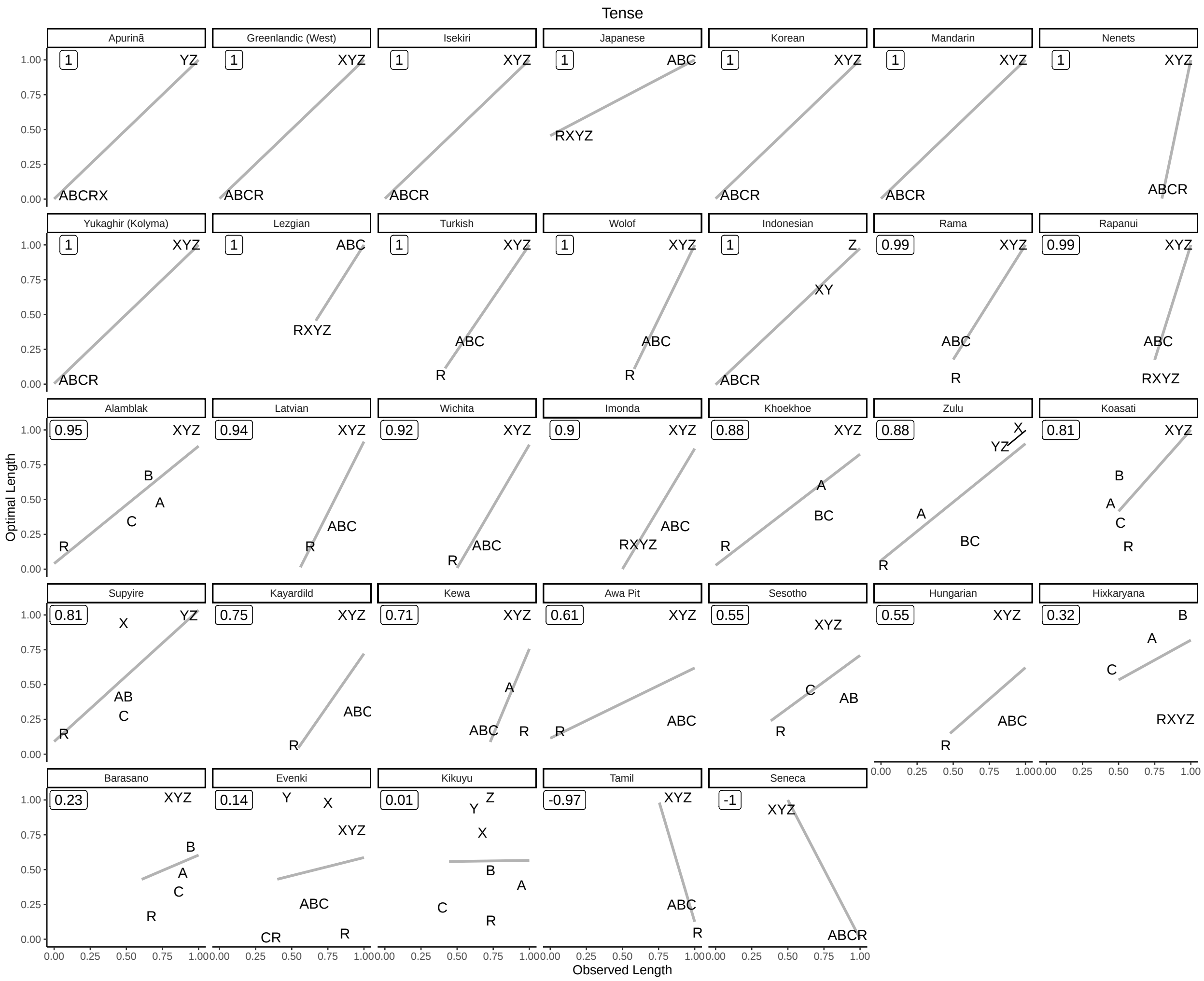

Fig. S7. Optimal lengths against observed lengths for individual languages in our tense data set. Correlations are shown at the top left of each panel.

Instead of a binary branching structure, Harbour (33, 34) explains the diversity of attested number systems using a feature based approach and lattice theory. Under this account, numerical systems can be described in terms of three binary features reflecting computations on a lattice. While this account covers attested data, it differs from our account in two important respects. First, Harbour's (34) account does not handle optional distinctions as Corbett's account and our account do. Second, indefinite feature values (e.g., paucal) must have their boundaries set by additional socio-cultural conventions; whereas, our account derives probable boundaries from functional pressures. In contrast to the accounts of both Corbett and Harbour, our framework provides a probability distribution over possible systems rather than a generative language for attested systems.

By traversing the Pareto frontier in Figure 2, we can identify systems with increasing numbers of feature values that achieve optimal trade-offs between complexity and information loss. The left column of Figure S9 shows stochastic systems that lie directly on the Pareto frontier, and the right column shows the deterministic systems that are closest to these stochastic systems. Each line in each panel shows a different feature value. For comparison, Figure S10 shows encoders represented in our sample of attested systems, and a full list of systems is provided in Table S7. In line with all prior accounts, the theoretically optimal two term system captures the SG/PL distinction. Our data, however, includes one two-term language that does not make this distinction. Japanese distinguishes between GENERAL and PL, and although this system is not strictly optimal it lies extremely close to the Pareto frontier.

For systems with three terms, the number hierarchy allows a distinction between SG, DU and PL, and the binary branching typology allows an additional system that distinguishes SG, PAUC, and PL. Our analysis suggests that the optimal three term system splits the plural into an indefinite feature value and, if languages had a pressure towards deterministic systems the closest one distinguishes SG, DU and PL. In our 10 state space, the optimal stochastic system would be consistent with either a SG/PAUC/PL split as seen in Bayso or a SG/PL/GPL[†] split as seen in Fula. That being said, if we scaled up our analysis to a larger state space, it is unlikely that the indeterminate categories would map to a PL/GPL distinction, suggesting that this attested distinction does not arise due to the tradeoff between information loss and complexity. The other attested three

[†] GPL stands for the greater plural, which corresponds to an exceptionally large number.

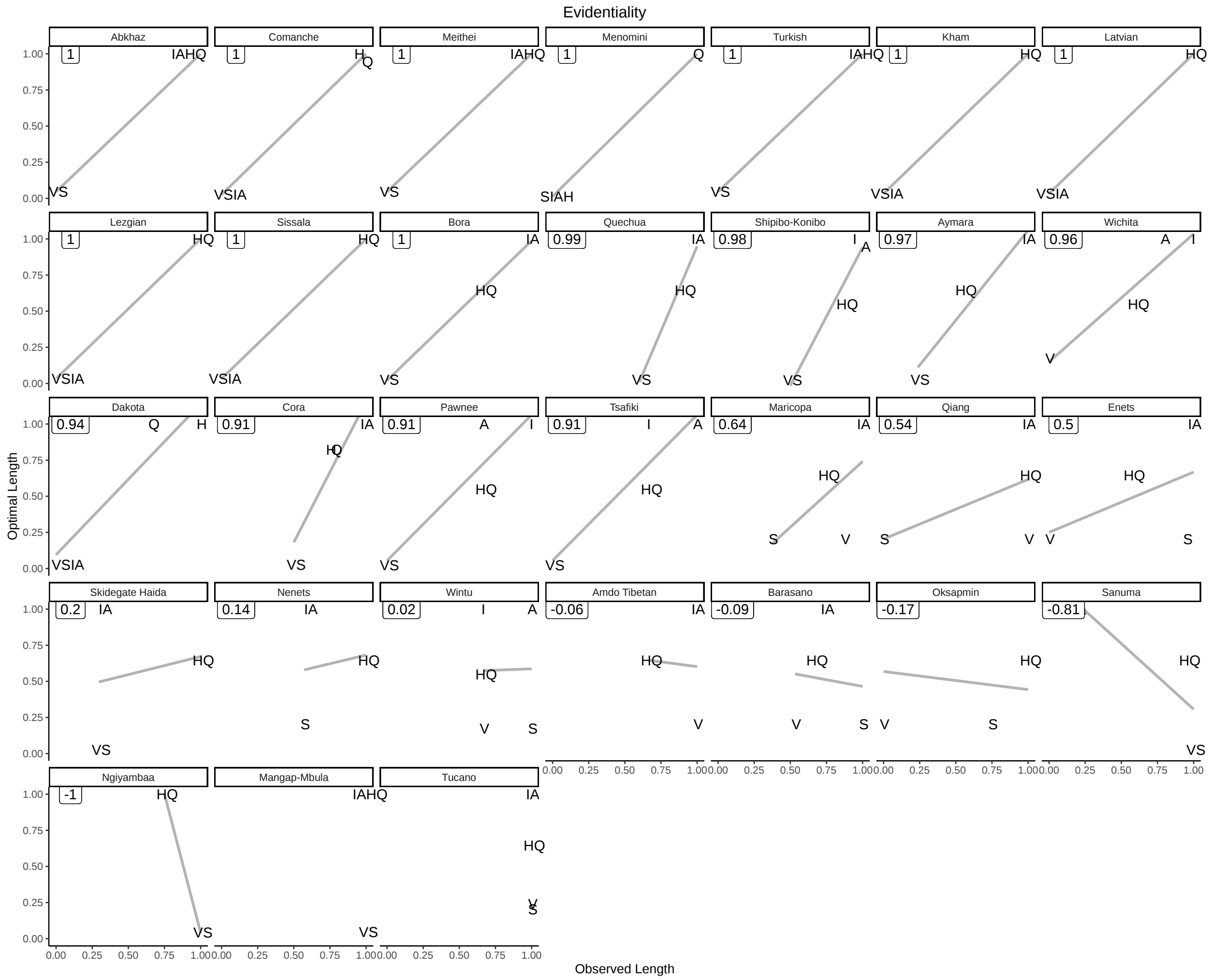

Fig. S8. Optimal lengths against observed lengths for individual languages in our evidentiality data set.

term system is SG/DU/PL, aligning with the near optimal deterministic system. Languages with optional dual like Slovene are closer to optimal than languages with non-optional dual like Sanskrit because collapsing across the optional dual results in the optimal two term system. That being said, languages with dual are still close to the optimal trade-off.

The optimal four term system according to our analysis is SG/DU/PAUC/PL, which is a very common system in Austronesian languages. There are a few other four term systems worth discussing. While making the same distinctions as the optimal system, both the dual and paucal in Longgu are optional. The resulting collapsed system would be the optimal two term system; however, the additional complexity of multiple optional terms pushes Longgu further away from the Pareto frontier. Similar to three term systems with two indeterminate categories, there are four term systems like Kaytetye which have a PL/GPL distinction instead of a PAUC/PL distinction. As for three term systems, it is again unclear whether four term systems with the PL/GPL distinction would continue to be near-optimal in an expanded state space. Other attested four term languages like Ngan'gityemerri split SG/DU/TR/PL, where the trial is always optional. In these cases, the collapsed system would not result in an optimal system. Larike makes the same split but with both an optional trial and an optional dual, resulting in an optimal collapsed system. Yet Larike is similar to Longgu in being pushed further from the frontier for having two optional distinctions. Recall, however, that all attested systems are still much closer to the Pareto frontier than the unattested systems in our analysis.

The optimal five term stochastic system splits another indeterminate feature value from plural SG/DU/PAUC/GPAUC/PL, which we see in Sursurunga. Even with three indeterminate categories, we still see at least one language that places a different range on those categories. Mele-Fila makes distinctions between SG/DU/PAUC/PL/GPL. Again it is unlikely this system would remain near-optimal if we scaled up our state space. While the optimal deterministic system in our analysis is SG/DU/TR/QUAD/PL, the second most similar system to optimal is SG/DU/TR/PAUC/PL which is exactly what we see in Lihir and Tangga. In our coding, we assumed that the trial in Lihir and Tangga is not optional. In truth, there seems to be a lack of data on the usage of the trial in these languages. Following the wider pattern of trials being optional, we know that Marshallese follows the optimal deterministic system but with optional dual, trial and paucal.

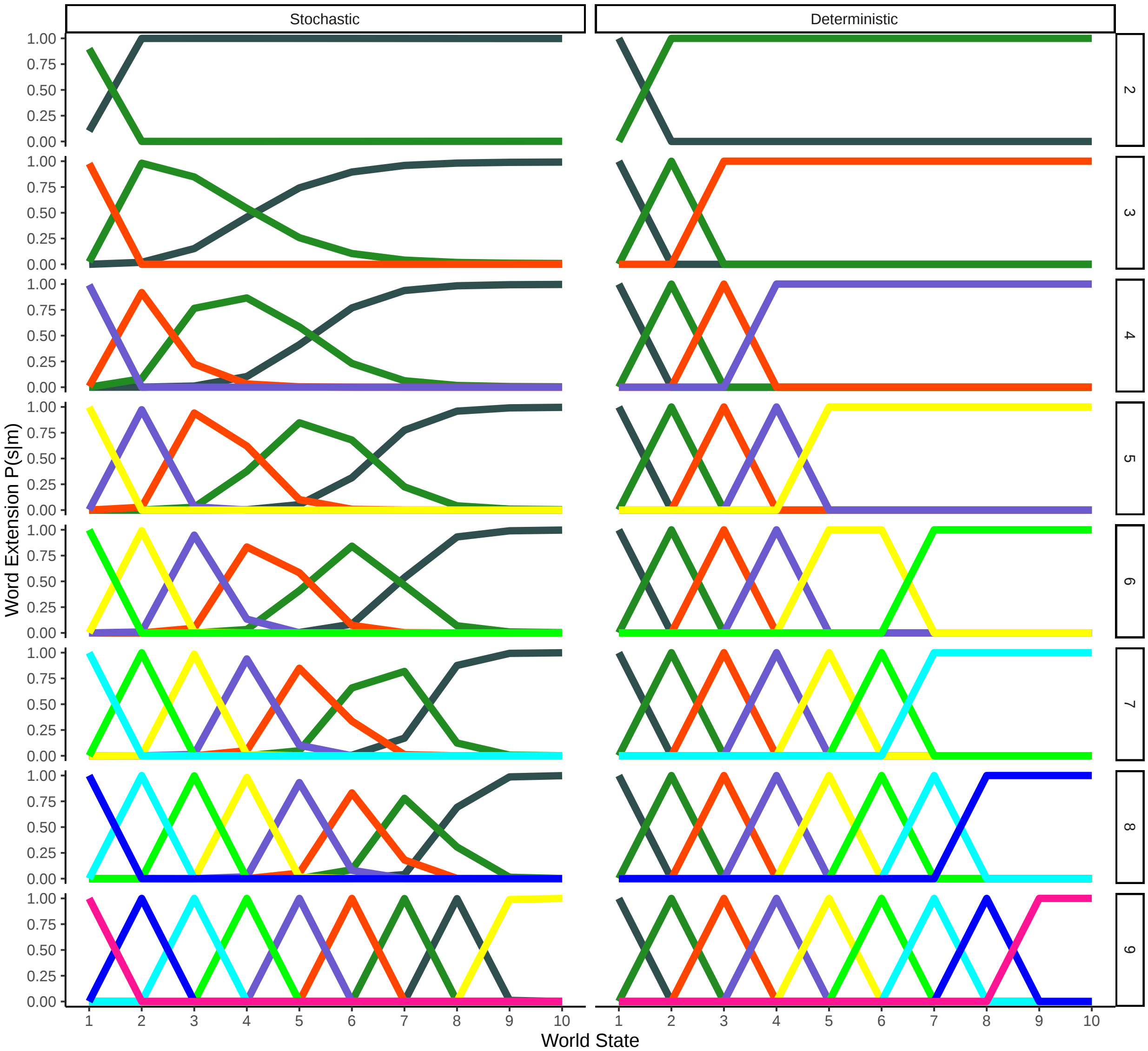

Fig. S9. Number Encoder Systems. Optimal stochastic systems and most similar deterministic systems according to the model.

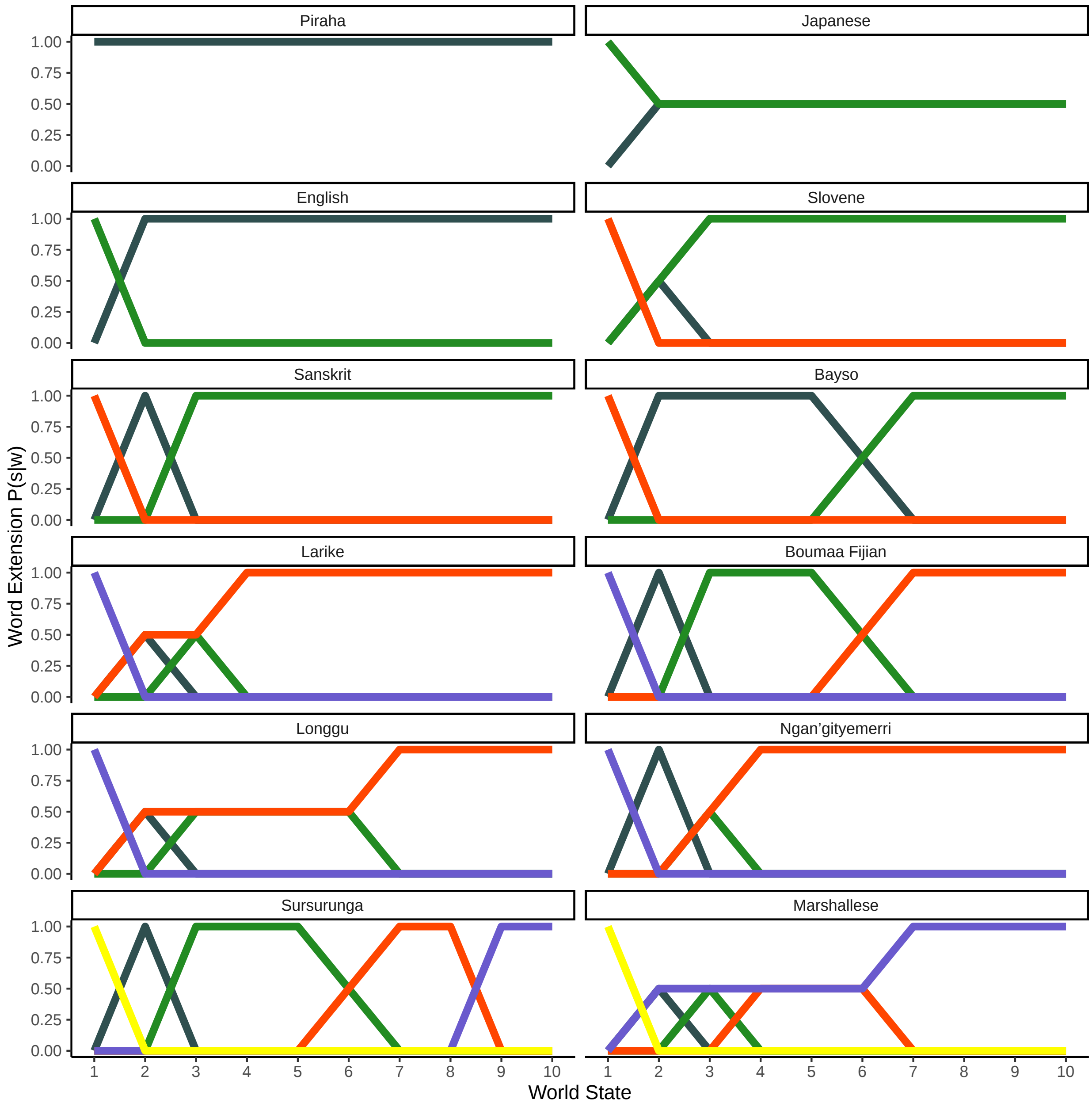

Fig. S10. Number Encoder Systems. A sample of attested language systems in our coding scheme.

There are a few typological insights we can take away from our analysis. If languages were only trying to optimize the trade-off between information loss and complexity: 1) they should initially introduce an indeterminate distinction before each new determinate one; 2) determinate distinctions should be introduced according to the number hierarchy; 3) if a language favors minimizing complexity, it should introduce an optional determinate instead of an indeterminate; 4) if a language favors minimizing information loss, it should introduce a determinate. It appears that having an optional determinate distinction is less complex than having an indeterminate distinction; however, an additional indeterminate distinction reduces information loss more than an optional determinate.

As far as we know, there are no six term systems; however, under our account, we would predict a six term system to split SG/DU/TR/PAUC/GPAUC/PL or SG/DU/TR/QUAD/PAUC/PL. Given the added complexity of a quadral, we should not be surprised to see the quadral (and possibly also the trial and dual) to be optional. Given the pattern in our data, it would not be surprising if the GPAUC/PL split matches ranges that would be better described as PL/GPL; however, this is not predicted by our model, which suggests that some other pressure drives the evolution of such systems.

B. Tense. Unlike number, we are not aware of any proposals for the "evolutionary" trajectory of tense systems. The majority of work on the grammaticalization of tense systems has focused on the evolution of the *forms* of tense markers, highlighting independent universal pathways for the grammatical forms (29). That being said, Comrie (35, pg. 85) points out the "general tendency of languages to have a better developed past than future system." Dahl (2, pg. 125) further elaborates on this observation:

> "If there is one or more distinctions of remoteness in a [Tense-Mood-Aspect] system, and reference can be made to objective time measures, one of the distinctions will be between "more than one day away" and "not more than one day away".

In our framework, we would then predict that the past tense would subdivide immediate past before remote past tense. We return to this prediction after describing the optimal system trajectory under our analysis.

The upper panels of Figure S11 show stochastic systems and their closest deterministic counterparts that emerge as we traverse the Pareto frontier from simple to complex. The bottom panels show the unique encoders in our set of attested tense systems. We provide a full list of attested systems in our analysis in Table S8. The optimal two term system splits past (ABC) from non-past (RXYZ). While the majority of the two term systems in our analysis like Apurinã make this distinction, the remainder make a distinction between future (XYZ) and non-future (ABCR) like Amharic. In a larger sample than ours, Velupillai (36) finds more two term systems with a future vs non-future split than we do. She argues that grammatical aspect reduces the ambiguity between past and present such that an explicit tense for future is more helpful for communication. While our model does not currently handle multiple dimensions, future work can extend our approach to explore this correlation. In our framework, if complexity is weighted more heavily than information loss then present should be grouped with past, but if information loss carries more weight then present should be grouped with future. Both attested systems, however, lie close to the Pareto frontier.

Looking at three term systems, the optimal system divides the timeline according to the higher level structure in our meaning space: past (ABC), present (R) and future (XYZ). This system is also the most frequently attested three term system in our analysis. There are two exceptions. Imonda splits the timeline into past (ABC), present (R) and non-past (ABCR), and can be viewed as having an optional present. Rapanui splits the timeline into past (ABC), not-past (RXYZ) and future (XYZ), and can be viewed as a having an optional future. As we have seen with grammatical number systems, optional distinctions reduce the complexity of a given n-term system relative to a non-optional system; however, they often collapse to the optimal $n-1$ term system. Imonda and Rapanui both collapse to the near optimal two term systems and, thus, still lie close to the Pareto frontier.

Scaling up to four term systems, both the stochastic and deterministic models predict a subdivision of the past tense, as seen in Khoekhoe and Asmat. Spanish and Diola-Fogny make optional distinctions for immediate and remote past respectively, and, thus collapse to the optimal three-term system and lie close to the optimal frontier. Interestingly, one language in our sample uses a four-term system that lies relatively far from the Pareto frontier. Whereas the rest of the three term systems distinguish between future, present and two past tenses, Hixkaryana distinguishes between three past tenses and a non-past feature value. Retaining a non-past feature value instead of splitting into present and future results in a substantial information loss.

The optimal five term system splits into two past tenses, a present tense and two future tenses. Similar to the four term systems, some languages split the past tense into remote and general past like Zulu; whereas, other languages split the past tense into general and recent past like Grebo. Both types of attested optimal system always split the future into near future and general future. One five term system in our sample, Evenki, makes the optimal three term system split augmented with an optional general past (B) and an optional immediate future (X). Collapsing to the optimal three term system, Evenki still lies close to the Pareto frontier. Similar to Hixkaryana, we also find five term systems like Alamblak that carve up the past into three categories; however, these systems also split present tense and future tense, which mitigates the loss of communicative precision observed in Hixkaryana. As a result, they lie relatively close to the Pareto frontier.

The optimal deterministic six term system has three past tenses, a present tense and two future tenses; whereas, the optimal stochastic six term system has three futures, two pasts and a present tense. There is one language with six terms in our sample, Kikuyu. Kikuyu makes the same distinctions as the optimal deterministic six term system.

Looking at the patterns as a whole, if there is a present tense, there must be either a past or future tense. According to deterministic optimal systems, there should always be more past tenses than future tenses, suggesting that if there is an equal

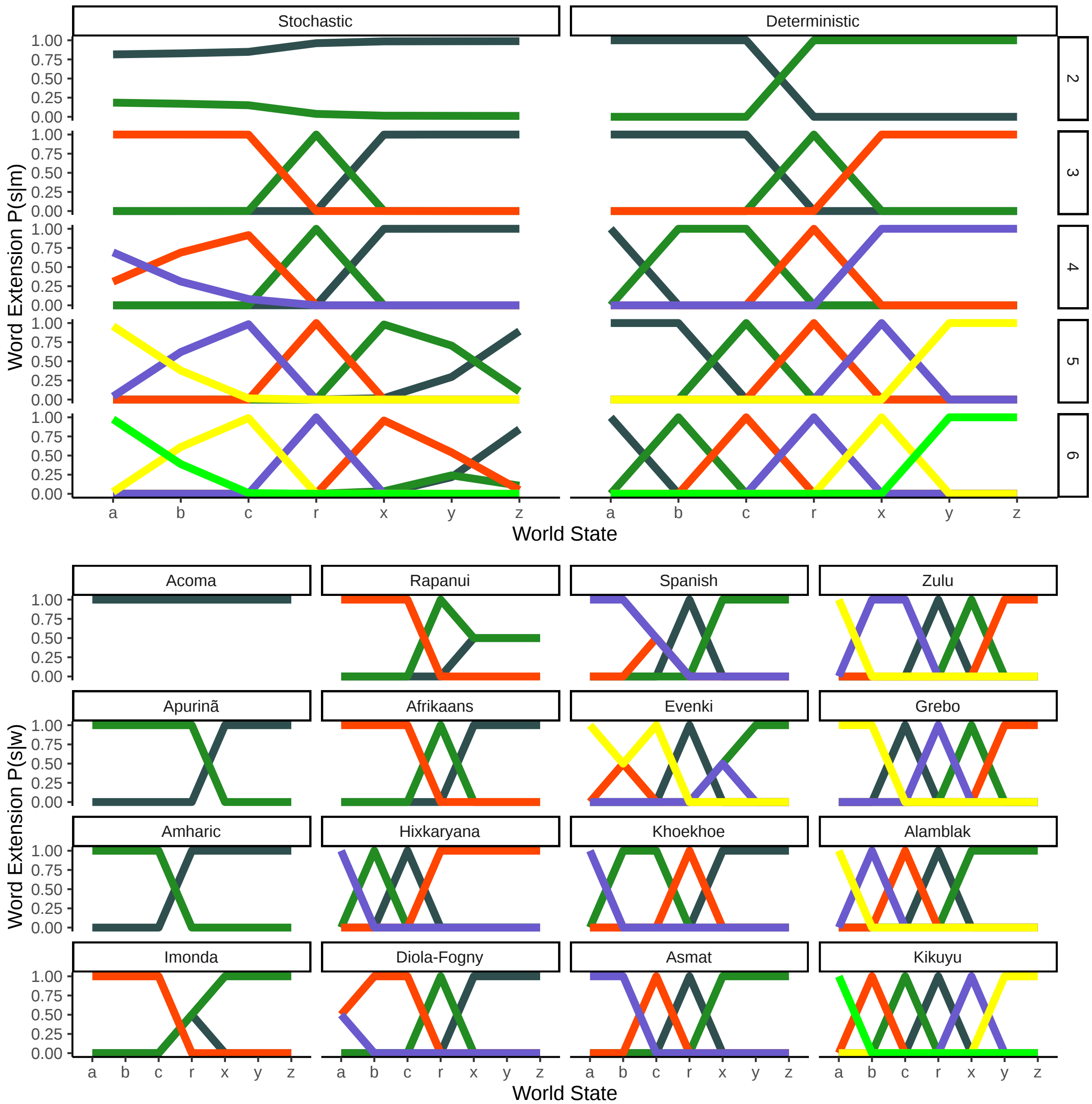

Fig. S11. Tense Encoder Systems. Top: Optimal stochastic systems and most similar deterministic systems according to the model. Bottom: A sample of attested language systems in our coding scheme.

number of past and future tenses, the next step should be to subdivide the past. Interestingly, optimal deterministic systems always split off the remote before the immediate; whereas, optimal stochastic systems divide remoteness relatively evenly. As a result, systems that split off the immediate before the remote also lie very close to the Pareto frontier. Although our theory does not recapitulate Dahl's observation that immediate past takes priority over remote past, our timeline is relatively coarse. Future data sets may allow need probabilities and temporal uncertainties to be estimated more precisely, which may allow a better alignment between the theory and Dahl's observation.

C. Evidentiality. To our knowledge, there are few claims about constraints on possible evidentiality systems, although Aikhenvald (3) has suggested that complex systems correlate with small population sizes. While number and tense come with clear underlying dimensions, our meanings for evidentiality are more speculative. We know that humans do not automatically tag their memories with information source and that source monitoring ability is relatively poor among Western populations (37, 38). The speaker distributions in Figure 1 are therefore fairly broad, although constrained by the hierarchy over states that we have assumed. Given our uncertainty about the underlying semantic representation, our analysis of evidentiality should be viewed as more provisional than our analyses of number and tense.

The upper panels of Figure S12 show stochastic systems and their closest deterministic counterparts that emerge as we traverse the Pareto frontier from simple to complex. The bottom panels show the unique encoders in our set of attested evidentiality systems. A complete list of attested systems in our data set is provided in Table S9.

The optimal two-term system according to our model distinguishes reported events (HQ) from non-reported events. This distinction is generalized in (3) to firsthand vs. non-firsthand. In attested systems, inferred and assumed sources can be grouped with firsthand as in Achagua or with non-firsthand as in Abkhaz. In practice, both systems lie close to the Pareto frontier. There are several attested two term systems that do not follow this pattern. Looking first at systems with non-firsthand categories, some systems like Chemehuevi include a more specific quotative feature value (Q) instead of picking out reported events more generally. A few Athabaskan systems (e.g., Chipewyan) construct their non-firsthand feature value from hearsay and inference. Both systems break the hierarchy over the meaning space assumed by our model. For systems like Chemehuevi, a quotative feature value should not exist without a hearsay feature value. For systems like Chipewyan, the non-firsthand feature value should not exclude assumed and quoted sources. That being said, they still lie near the Pareto frontier because the wider non-firsthand feature value (IAHQ) has much lower need probability than the narrow firsthand feature value (VS).

Looking at attested systems with firsthand categories, some systems like Mansi narrow their firsthand feature value to only include visual information sources. Similarly, Meithei narrows its firsthand feature value to only include sensory information sources. In contrast to the non-firsthand narrowed systems, these systems lie quite far from the Pareto frontier because they break the meaning hierarchy where the need probability is high. A close look at these systems suggests that the meanings they convey are not adequately captured by the underlying evidentiality dimension assumed by our model. For Mansi, Northern Kanthy, Svan, and Turkish, evidentiality correlates with mirativity—i.e., a marker expressing surprise. For example, even after watching Captain America wield Thor's hammer, one can say "Cap can wield Mjölnir" with the non-firsthand marker because only Thor was thought to be capable of using the hammer, making Captain America's use surprising. Similarly, evidentiality in Nenets has a complex interaction with the mood system. In Meithei, evidentials interact with person, such that the non-firsthand marker can apply when the listener, rather than the speaker, lacks firsthand knowledge. For example, one can say "I have eaten" with a non-firsthand marker even though the speaker has firsthand knowledge of having eaten, because the listener does not know. We do not know enough about Megrelian and Xakas to discuss usage in these languages, but if we speculate based on geography or phylogeny, we would expect Megrelian and Xakas to behave like Svan and Turkish. While we appeal to correlations with other dimensions to explain these sub-optimal systems, correlations likely exist with other dimensions for all evidentiality systems, and for all other grammatical systems as well. Future analyses will therefore need to scale up to meaning spaces with multiple semantic dimensions in order to draw more definitive conclusions about constraints on linguistic diversity and the optimality of complete grammatical systems.

Moving on to three term systems, the optimal system in our model distinguishes between direct evidence (VS), indirect evidence (IA) and reported evidence (HQ), again borrowing Aikhenvald's (3) terminology. The majority of three term systems, including Amdo Tibetan and Quechua, follow the optimal system. A few languages do not make the split along the top level of our meaning hierarchy and instead further split a reported evidential into two categories. For example, Comanche has a general quotative marker and adopts a marker reflecting information that the speaker does not know in narrative discourse. The last three term system in our analysis, Siona, splits visual, sensory and non-firsthand sources. Similar to Mansi, our analysis suggests that splitting visual and sensory categories at this stage would be sub-optimal. However, closer analysis of Siona shows that evidentiality correlates with the degree of control the speaker had over the situation. Full control receives visual marking, partial control receives sensory marking and no control receives non-firsthand marking. While these systems do not make all top level distinctions in the meaning hierarchy, the distinctions they implement still respect the hierarchy and so they do lie relatively close to the Pareto frontier.

As meanings are not strongly biased within each hierarchical cluster, the optimal stochastic systems do not make strong predictions for four and five term systems. The optimal four term system weakly subdivides the indirect evidence feature value into assumed and inferred sources, as seen deterministically in Lillooet. Other attested systems, such as Cora, subdivide the reported feature value into hearsay and quotative sources instead. Others, such as Barasano, subdivide the direct evidence feature value into visual and sensory sources. These systems all respect the meaning hierarchy and, thus, lie near the Pareto frontier. One language, Kwakiutl, has firsthand (VSI), assumed (A), hearsay (H) and quotative (Q) markers, which violates the assumed meaning hierarchy. This violation, however, occurs in a part of the hierarchy with low need probability, meaning that

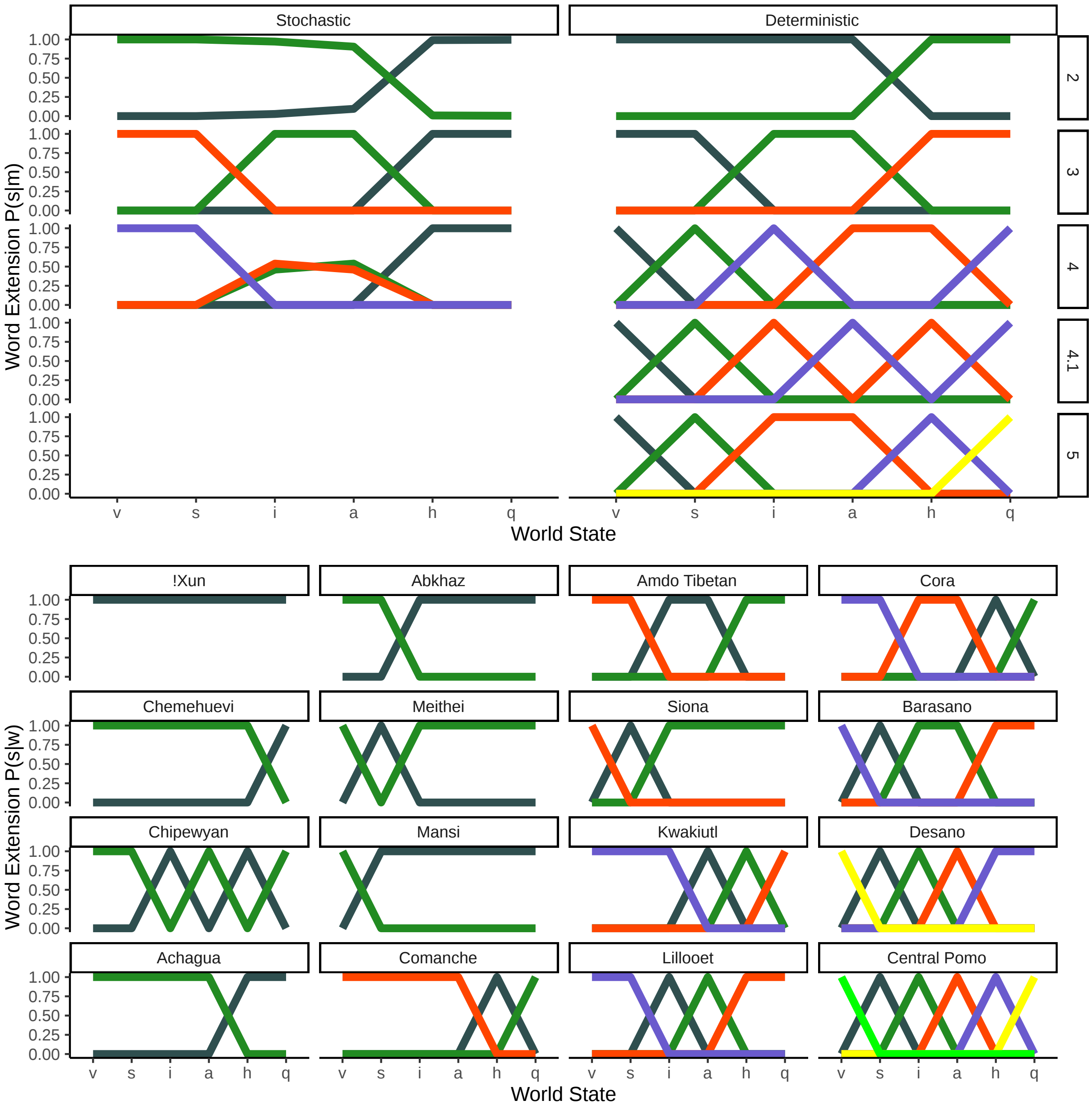

Fig. S12. Evidentiality Encoder Systems. Top: Optimal stochastic systems and most similar deterministic systems according to the model. Bottom: A sample of attested language systems in our coding scheme.

the system does not lie too far from the Pareto frontier.

The bias in meanings and relative frequency between visual/sensory and hearsay/quotative are mathematically equivalent. Both distinctions therefore emerge at the same rate, and there is no optimal five term system. The five term systems in our sample (e.g., Desano) fully subdivide direct and indirect evidence while leaving the reported evidence intact. Again this system obeys the meaning hierarchy and, therefore, lies along the Pareto frontier.

The overall pattern with our evidentiality analysis is that distance to the Pareto frontier corresponds to how many distinctions in the meaning hierarchy are preserved. This finding corresponds nicely with Gentner & Bowerman (39)'s proposal that the naturalness of a concept corresponds to its typological frequency. Following the meaning hierarchy, optimal systems divide the top level clusters before making finer distinctions. When choosing to divide hierarchical clusters at the same level, the model predicts dividing the feature value with the lowest need probability. This prediction is consistent with Dahl's (2) intuition that all else being equal, less frequent categories should be preferred in use over a broad frequent feature value; otherwise, the less frequent feature value would never be used at all.

Language (Source)	System	Complexity	Info. Loss	Frontier Dist.	gNID	Tradeoff
Piraha (1—Corbett, 2000: 50-51)	GENERAL	0.00	1.44	0.00	0.00	1.02
Kawi (1—Corbett, 2000: 51)	GENERAL	0.00	1.44	0.00	0.00	1.02
Japanese (1—Corbett, 2000: 14)	GENERAL/PL	0.32	1.14	0.01	-0.24	1.03
English (1—Corbett, 2000: 9)	SG/PL	0.94	0.55	0.01	0.06	1.16
Russian (1—Corbett, 2000: 4)	SG/PL	0.94	0.55	0.01	0.06	1.16
Slovene (1—Corbett, 2000: 43-45)	SG/DU_O/PL	1.05	0.45	0.01	0.03	1.17
Bayso (1—Corbett, 2000: 22)	SG/PAUC/PL	1.12	0.40	0.01	0.06	1.24
Avar (1—Corbett, 2000: 22)	SG/PAUC/PL	1.12	0.40	0.01	0.06	1.24
Banyun (1—Corbett, 2000: 31)	SG/PL/GPL	1.12	0.40	0.01	0.06	1.24
Fula (1—Corbett, 2000: 31)	SG/PL/GPL	1.12	0.40	0.01	0.06	1.24
Hamer (1—Corbett, 2000: 32-33)	SG/PL/GPL	1.12	0.40	0.01	0.06	1.24
Longgu (1—Corbett, 2000: 46-48)	SG/DU_O/$PAUC_O$/PL	1.13	0.40	0.02	0.05	1.20
Larike (1—Corbett, 2000: 21)	SG/DU_O/TR_O/PL	1.13	0.40	0.03	0.04	1.19
Marshallese (1—Corbett, 2000: 29-30)	SG/DU_O/TR_O/$PAUC_O$/PL	1.21	0.35	0.04	0.05	1.23
Upper Sorbian (1—Corbett, 2000: 20)	SG/DU/PL	1.29	0.26	0.02	0.05	1.47
Sanskrit (1—Corbett, 2000: 45)	SG/DU/PL	1.29	0.26	0.02	0.05	1.47
Ngan'gityemerri (1—Corbett, 2000: 21-22;43-44)	SG/DU/TR_O/PL	1.35	0.22	0.02	0.03	1.50
Marrithiyel (1—Corbett, 2000: 22;43)	SG/DU/TR_O/PL	1.35	0.22	0.02	0.03	1.50
Anindilyakwa (1—Corbett, 2000: 22;43)	SG/DU/TR_O/PL	1.35	0.22	0.02	0.03	1.50
Yimas (1—Corbett, 2000: 23)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Meryam Mir (1—Corbett, 2000: 23)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Boumaa Fijian (1—Corbett, 2000: 23-24)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Paamese (1—Corbett, 2000: 24)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Manam (1—Corbett, 2000: 24)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Ambrym (1—Corbett, 2000: 24)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Kwaio (1—Corbett, 2000: 25)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Sa'a (1—Corbett, 2000: 25)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Langalanga (1—Corbett, 2000: 25)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Lau (1—Corbett, 2000: 25)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Ungarinjin (1—Corbett, 2000: 25)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Murrinh-Patha (1—Corbett, 2000: 25)	SG/DU/PAUC/PL	1.43	0.16	0.01	0.02	1.71
Kaytetye (1—Corbett, 2000: 33)	SG/DU/PL/GPL	1.43	0.16	0.01	0.02	1.71
Mokilese (1—Corbett, 2000: 34)	SG/DU/PL/GPL	1.43	0.16	0.01	0.02	1.71
Sursurunga (1—Corbett, 2000: 26-27)	SG/DU/PAUC/GPAUC/PL	1.47	0.14	0.02	0.04	1.84
Mele-Fila (1—Corbett, 2000: 35)	SG/DU/PAUC/PL/GPL	1.47	0.14	0.02	0.04	1.84
Lihir (1—Corbett, 2000: 25-26)	SG/DU/TR/PAUC/PL	1.57	0.07	0.01	0.01	2.37
Tangga (1—Corbett, 2000: 29)	SG/DU/TR/PAUC/PL	1.57	0.07	0.01	0.01	2.37

Table S7. Attested grammatical number systems and their parameters according to the model. The trade-off parameter, β, provided corresponds to the beta value that maximizes similarity. SG-singular; PL-plural; DU-dual; TR-trial; PAUC-paucal; GPAUC-greater paucal; GPL-greater plural. $_O$ denotes optional.

Language (Source)	System	Complexity	Info. Loss	Frontier Dist.	gNID	Tradeoff
Acoma (40—Miller, 1965)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Ainu (41—Refsing, 1986)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Amele (42—Roberts, 1987: 227-228)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Araona (43—Pitman, 1980)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Arapesh (Mountain; 44—Conrad & Wogiga, 1991)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Bagirmi (45—Stevenson, 1969)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Berber (Middle Atlas; 46—Penchoen, 1973)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Cebuano (2—Dahl, 1985: 160)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Cree (Plains; 47—Wolfart, 1996)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Dani (Lower Grand Valley; 48—Bromley, 1981: 61-65)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Garífuna (29—Bybee et al., 1994)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Hausa (49—Newman, 2000)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Hmong Njua (50—Harriehausen, 1990)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Jakaltek (51—Grinevald Craig, 1977)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Ju|'hoan (52—Dickens, 1991)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Karok (29—Bybee et al., 1994)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Ket (53—Krejnovich, 1968)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Khmer (54—Jacob, 1968)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Kiowa (55—Watkins, 1984)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Koyraboro Senni (56—Heath, 1999)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Kutenai (57—Canestrelli & Boas, 1927)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Lahu (29—Bybee et al., 1994)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Lakhota (58—Buechel, 1939)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Lango (59—Noonan, 1992)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Lavukaleve (60—Terrill, 1999)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Luganda (61—WALS)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Luvale (62—Horton, 1949)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Malagasy (63—Dez, 1980)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Mapudungun (64—Smeets, 1989)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Maricopa (65—Gordon, 1986)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Maybrat (66—Dol, 1999)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Meithei (67—Chelliah, 1997)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Motu (29—Bybee et al., 1994)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Ngambay (29—Bybee et al., 1994)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Nicobarese (Car; 29—Bybee et al. 1994)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Paiwan (68—Chang, 2006)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Pirahã (69—Everett, 1986)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Rendille (61—WALS)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Sango (70—Thornell, 1997)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Sanuma (71—Borgman, 1990: 167)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Semelai (72—Kruspe, 2004)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Suena (73—Wilson, 1974: 39-40)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Tagalog (74—Schachter & Otanes, 1972)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Tukang Besi (75—Donohue, 1999)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Vietnamese (76—Nguyen, 1997)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Yagua (77—Payne & Payne, 1990: 386-87)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Yucatec (78—Bohnemeyer, 2002)	(ABCRXYZ)	0.00	1.01	0.00	0.00	1.29
Apurinã (79—Facundes, 2000: 517)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Canela-Krahô (80—Popjes & Popjes, 1986)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Chamorro (81—Topping, 1973)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Greenlandic (West; 2—Dahl, 1985: 165)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Indonesian (2—Dahl, 1985: 160)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Isekiri (2—Dahl, 1985: 178)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Javanese (2—Dahl, 1985: 161)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Kayardild (82—Evans, 1995)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Khmu' (2—Dahl, 1985: 174)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Lai (29—Bybee et al., 1994)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Mandarin (2—Dahl, 1985: 180)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Nivkh (83—Panfilov, 1962)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Oneida (2—Dahl, 1985: 173)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Seneca (2—Dahl, 1985: 173)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Thai (2—Dahl, 1985: 174)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Wari' (84—Everett & Kern, 1997)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Yoruba (2—Dahl, 1985: 179)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Yukaghir (Kolyma; 85—Maslova, 1999)	(ABCR)(XYZ)	0.81	0.50	0.04	0.27	1.80
Amharic (2—Dahl, 1985: 156)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Arabic (Egyptian; 86—Eisele, 1999)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Arabic (Tunisian; 2—Dahl, 1985: 155)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Avokaya (61—WALS)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Awa Pit (87—Curnow, 1997: 176)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Brahui (88—Andronov, 1980)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Burushaski (89—Berger, 1998: 142, 160-161)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Daga (90—Murane, 1974: 53)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82

Fijian (91—Dixon, 1988: 73)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Finnish (2—Dahl, 1985: 181)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Inuktitut (Salluit; 92—Hayashi & Spreng, 2005)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Japanese (2—Dahl, 1985: 158)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Khalkha (93—Svantesson, 1991)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Korean (94—Lee, 1991: 225)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Lezgian (95—Haspelmath, 1993: 146)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Mangarrayi (96—Merlan, 1982: 61)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Maori (2—Dahl, 1985: 163)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Martuthunira (97—Dench, 1995: 144)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Nenets (98—Salminen, 1997)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Oromo (Harar; 99—Owens, 1985: 73)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Rama (100—Grinevald, 1990)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Sundanese (2—Dahl, 1985: 139)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Tiwi (101—Osborne, 1974: 40)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Warao (102—Romero-Figueroa, 1997: 95)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Yaqui (103—Dedrick & Casad, 1999: 318)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Zoque (Copainalá; 104—Harrison et al., 1981: 92)	(ABC)(RXYZ)	0.85	0.47	0.04	0.25	1.82
Imonda (105—Seiler, 1985)	(ABC)(R)(RXYZ)	1.03	0.38	0.06	0.10	1.85
Rapanui (106—Du Feu, 1996: 156)	(ABC)(RXYZ)(XYZ)	1.08	0.32	0.03	0.09	1.88
Hixkaryana (107—Derbyshire, 1979: 138-139)	(A)(B)(C)(RXYZ)	1.26	0.42	0.21	0.38	2.36
Afrikaans (2—Dahl, 1985: 166)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Akan (2—Dahl, 1985: 179)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Alawa (2—Dahl, 1985: 159)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Armenian (Eastern; 108—Dum-Tragut, 2009)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Azerbaijani (2—Dahl, 1985: 159-160)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Bandjalang (2—Dahl, 1985: 159)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Beja (2—Dahl, 1985: 154-155)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Bengali (2—Dahl, 1985: 167)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Bulgarian (2—Dahl, 1985: 172)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Chukchi (109—Dunn, 1999)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Didinga (110—Lohitare et al., 2012)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
English (2—Dahl, 1985: 166)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
French (2—Dahl, 1985: 170)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Georgian (2—Dahl, 1985: 164)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
German (2—Dahl, 1985: 166)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Gooniyandi (111—McGregor, 1990: 20)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Greek (Modern; 2—Dahl, 1985: 169)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Guaraní (2—Dahl, 1985: 159)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Hebrew (Modern; 2—Dahl, 1985: 114)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Hindi (2—Dahl, 1985: 168-169)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Hungarian (2—Dahl, 1985: 181)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Icelandic (112—Svenonius, 2002)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Kannada (113—Sridhar, 1990: 219)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Kriol (Fitzroy Crossing; 2—Dahl, 1985: 166)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Kurmanji (2—Dahl, 1985: 168)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Latvian (114—Hewson & Bubenik, 1987)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Makah (115—Davidson, 2002)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Maltese (2—Dahl, 1985: 156)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Mixtec (Chalcatongo; 116—Macaulay, 1996)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Montagnais (117—Drapeau, 2014)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Ngiyambaa (118—Donaldson, 1980: 160)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Punjabi (2—Dahl, 1985: 168)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Persian (2—Dahl, 1985: 168)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Portuguese (2—Dahl, 1985: 171-172)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Quechua (Cochabamba; 2—Dahl 1985: 158)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Quechua (Imbabura; 2—Dahl 1985: 144)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Romanian (2—Dahl, 1985: 171-172)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Russian (2—Dahl, 1985: 172)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Somali (119—Saeed, 1999)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Swahili (120—Lindfors, 2003)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Swedish (2—Dahl, 1985: 166)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Tamil (2—Dahl, 1985: 165-166)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Tenyer (2—Dahl, 1985: 177)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Tigrinya (2—Dahl, 1985: 156)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Turkish (2—Dahl, 1985: 157)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Wichita (121—Rood, 1976)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Wolaytta (122—Lamberti & Sottile, 1997)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Wolof (2—Dahl, 1985: 180)	(ABC)(R)(XYZ)	1.52	0.08	0.00	0.00	4.95
Diola-Fogny (123—Sapir, 1965: 30-35)	(A)(ABC)(R)(XYZ)	1.61	0.07	0.00	0.00	5.53
Kewa (124—Franklin, 1971)	(A)(ABC)(R)(XYZ)	1.61	0.07	0.00	0.00	5.53
Spanish (2—Dahl, 1985: 171)	(ABC)(C)(R)(XYZ)	1.61	0.07	0.00	0.00	5.41
Evenki (125—Nedjalkov, 1997: 235)	(ABC)(B)(R)(X)(XYZ)	1.68	0.07	0.02	0.02	5.84
Khoekhoe (126—Hagman, 1977: 64)	(A)(BC)(R)(XYZ)	1.76	0.05	0.01	0.05	10.40

Asmat (127—Voorhoeve, 1965: 98)	(AB)(C)(R)(XYZ)	1.80	0.05	0.01	0.06	11.03
Sesotho (2—Dahl, 1985: 173)	(AB)(C)(R)(XYZ)	1.80	0.05	0.01	0.06	11.03
Alamblak (128—Bruce, 1984: 133)	(A)(B)(C)(R)(XYZ)	1.93	0.04	0.01	0.06	13.19
Barasano (129—Jones & Jones, 1991: 84-86)	(A)(B)(C)(R)(XYZ)	1.93	0.04	0.01	0.06	13.19
Koasati (130—Kimball, 1991: 203)	(A)(B)(C)(R)(XYZ)	1.93	0.04	0.01	0.06	13.19
Wichí (131—Wdzenczny)	(A)(B)(C)(R)(XYZ)	1.93	0.04	0.01	0.06	13.19
Zulu (2—Dahl, 1985: 175)	(A)(BC)(R)(X)(YZ)	2.01	0.02	0.00	0.05	14.61
Grebo (132—Innes, 1966: 54-87)	(AB)(C)(R)(X)(YZ)	2.05	0.02	0.00	0.05	15.45
Supyire (133—Carlson, 1994: 329-334)	(AB)(C)(R)(X)(YZ)	2.05	0.02	0.00	0.05	15.45
Kikuyu (2—Dahl, 1985: 175)	(A)(B)(C)(R)(X)(YZ)	2.18	0.01	0.00	0.03	19.66

Table S8. Attested tense systems and their parameters according to the model. A-remote past; B-past; C-near past; R-present; X-immediate future; Y-future; Z-remote future.

Language	System	Complexity	Info. Loss	Frontier Dist.	gNID	Beta
!Xun (3—Aikhenvald, 2004: 149)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Adioukrou (3—Aikhenvald 2004: 135)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Afghan Persian (3—Aikhenvald 2004: 109)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Aguaruna (3—Aikhenvald 2004: 137)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Andean Spanish (3—Aikhenvald 2004: 21)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Arabic (3—Aikhenvald 2004: 10)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Avar (3—Aikhenvald 2004: 38)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Chukchi (3—Aikhenvald 2004)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Colombian Spanish (3—Aikhenvald 2004: 141)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Dani (3—Aikhenvald 2004)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Donno Saw (3—Aikhenvald 2004: 133)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Dutch (3—Aikhenvald 2004: 17)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Dyirbal (3—Aikhenvald 2004: 132)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
English (3—Aikhenvald 2004: 10)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
French (3—Aikhenvald 2004: 17)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Georgian (3—Aikhenvald 2004: 38)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
German (3—Aikhenvald 2004)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Greek (3—Aikhenvald 2004: 150)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Hebrew (3—Aikhenvald 2004: 10)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Irantxe (3—Aikhenvald 2004: 75)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Ishkashim (3—Aikhenvald 2004: 38)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Istanbul Judezmo (3—Aikhenvald 2004: 114)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Japanese (3—Aikhenvald 2004: 81)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Kambera (3—Aikhenvald 2004: 140-141)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Ket (3—Aikhenvald 2004: 290)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Korean (3—Aikhenvald 2004)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Makah (3—Aikhenvald 2004)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Nivkh (3—Aikhenvald 2004: 290)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Panare (3—Aikhenvald 2004: 253)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Paumari (3—Aikhenvald 2004: 284)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Persian (3—Aikhenvald 2004: 38)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Portuguese (3—Aikhenvald 2004: 37)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Russian (3—Aikhenvald 2004)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Santali (3—Aikhenvald 2004: 131)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Spanish (3—Aikhenvald 2004)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Tuvaluan (3—Aikhenvald 2004: 359)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Ubykh (3—Aikhenvald 2004: 293)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Vlach Romani (3—Aikhenvald 2004: 38)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Wasco-Wishram (3—Aikhenvald 2004: 39)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Weyewa (3—Aikhenvald 2004: 359)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Zazaki (3—Aikhenvald 2004: 38-39)	(VSIAHQ)	0.00	0.38	0.00	1.00	1.05
Agul (3—Aikhenvald 2004: 143)	(VSIAH)(Q)	0.16	0.27	0.03	0.56	1.21
Chemehuevi (3—Aikhenvald 2004: 51)	(VSIAH)(Q)	0.16	0.27	0.03	0.56	1.21
Menomini (3—Aikhenvald 2004: 33)	(VSIAH)(Q)	0.16	0.27	0.03	0.56	1.21
Chipewyan (3—Aikhenvald 2004: 31)	(VSAQ)(IH)	0.21	0.27	0.06	0.36	1.30
Hare (3—Aikhenvald 2004: 31)	(VSAQ)(IH)	0.21	0.27	0.06	0.36	1.30
Kato (3—Aikhenvald 2004: 31)	(VSAQ)(IH)	0.21	0.27	0.06	0.36	1.30
Achagua (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Arabela (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Arrernte (3—Aikhenvald 2004: 33)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Baniwa (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Cashibo (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Cupeno (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Daw (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Enga (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Estonian (3—Aikhenvald 2004: 33)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Gaviao (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Guarani, Paraguayan (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Ignaciano (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Karitiana (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Kham (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Kombai (3—Aikhenvald 2004: 137)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Latvian (3—Aikhenvald 2004: 33)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Lezgian (3—Aikhenvald 2004: 31)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Livonian (3—Aikhenvald 2004: 33)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Omaha (3—Aikhenvald 2004: 33)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Pareci (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Piapoco (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Piro (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Potawatomi (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Resigaro (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Shoshone (3—Aikhenvald 2004: 33-34)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Sissala (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99

Surui (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Tauya (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Terena (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Warlpiri (3—Aikhenvald 2004: 62)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Waura (3—Aikhenvald 2004: 32)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Yankunytjatjara (3—Aikhenvald 2004: 33)	(VSIA)(HQ)	0.28	0.15	0.00	0.97	1.99
Comanche (3—Aikhenvald 2004: 50)	(VSIA)(H)(Q)	0.32	0.14	0.02	0.84	2.29
Dakota (3—Aikhenvald 2004: 51)	(VSIA)(H)(Q)	0.32	0.14	0.02	0.84	2.29
Tonkawa (3—Aikhenvald 2004: 51)	(VSIA)(H)(Q)	0.32	0.14	0.02	0.84	2.29
Abkhaz (3—Aikhenvald 2004: 29-30)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Albanian (3—Aikhenvald 2004: 40)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Arizona Tewa (3—Aikhenvald 2004: 42)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Armenian (3—Aikhenvald 2004: 39)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Bagvalal (3—Aikhenvald 2004: 155)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Balkan Slavic (3—Aikhenvald 2004: 40;158)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Bulgarian (3—Aikhenvald 2004: 288)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Cherokee (3—Aikhenvald 2004: 26-7)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Circassian (3—Aikhenvald 2004: 293)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Cree (3—Aikhenvald 2004: 42)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Cree/Montagnais/Naskapi (3—Aikhenvald 2004: 42)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Godoberi (3—Aikhenvald 2004: 28)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Hunzib (3—Aikhenvald 2004: 29)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Kalasha (3—Aikhenvald 2004: 28)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Kazakh (3—Aikhenvald 2004: 40)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Khowar (3—Aikhenvald 2004: 24)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Komi (3—Aikhenvald 2004: 28)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Kurdish (3—Aikhenvald 2004: 289)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Macedonian (3—Aikhenvald 2004: 40)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Mangap-Mbula (3—Aikhenvald 2004: 28)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Mari (3—Aikhenvald 2004: 28)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Megleno-Romanian (3—Aikhenvald 2004: 288)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Nepali (3—Aikhenvald 2004: 291)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Secoya (3—Aikhenvald 2004: 28)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Serbian (3—Aikhenvald 2004: 288)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Tajik (3—Aikhenvald 2004: 289)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Takelma (3—Aikhenvald 2004: 41)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Tarascan (3—Aikhenvald 2004: 41-42)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Tibetan (3—Aikhenvald 2004: 28)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Tsez (3—Aikhenvald 2004: 28)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Turkmen (3—Aikhenvald 2004: 40)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Udmurt (3—Aikhenvald 2004: 28)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Uyghur (3—Aikhenvald 2004: 40)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Uzbek (3—Aikhenvald 2004: 40)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Yanam (3—Aikhenvald 2004: 18)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Yukaghir (3—Aikhenvald 2004: 24;27)	(VS)(IAHQ)	0.36	0.12	0.01	0.91	2.57
Kwakiutl (3—Aikhenvald 2004: 59)	(VSI)(A)(H)(Q)	0.41	0.11	0.02	0.83	3.40
Amdo Tibetan (3—Aikhenvald 2004: 45)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Archi (3—Aikhenvald 2004: 29)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Aymara (3—Aikhenvald 2004: 43)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Bora (3—Aikhenvald 2004: 44)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Capanawa (3—Aikhenvald 2004: 46)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Diyari (3—Aikhenvald 2004: 34)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Hopi (3—Aikhenvald 2004: 45)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Jarawara (3—Aikhenvald 2004: 23-4;26)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Koreguaje (3—Aikhenvald 2004: 44)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Maidu (3—Aikhenvald 2004: 46)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Moseten (3—Aikhenvald 2004: 44)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Ngiyambaa (3—Aikhenvald 2004: 34)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Ponca (3—Aikhenvald 2004: 33-34)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Quechua (3—Aikhenvald 2004: 43)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Retuara (3—Aikhenvald 2004: 49)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Sanuma (3—Aikhenvald 2004: 18;46)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Shasta (3—Aikhenvald 2004: 43)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Shilluk (3—Aikhenvald 2004: 43)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Skidegate Haida (3—Aikhenvald 2004: 46)	(VS)(IA)(HQ)	0.42	0.08	0.00	1.00	8.79
Lillooet (3—Aikhenvald 2004: 59)	(VS)(I)(A)(HQ)	0.44	0.08	0.00	0.96	9.54
Shipibo-Konibo (3—Aikhenvald 2004: 55)	(VS)(I)(A)(HQ)	0.44	0.08	0.00	0.96	9.54
Thompson (3—Aikhenvald 2004: 59)	(VS)(I)(A)(HQ)	0.44	0.08	0.00	0.96	9.54
Cora (3—Aikhenvald 2004: 57)	(VS)(IA)(H)(Q)	0.47	0.08	0.00	0.92	9.86
Northern Embera (3—Aikhenvald 2004: 58)	(VS)(IA)(H)(Q)	0.47	0.08	0.00	0.92	9.86
Southeastern Tepehuan (3—Aikhenvald 2004: 58)	(VS)(IA)(H)(Q)	0.47	0.08	0.00	0.92	9.86
Mansi (3—Aikhenvald 2004: 31)	(V)(SIAHQ)	1.00	0.26	0.23	0.75	14.70
Megrelian (3—Aikhenvald 2004: 31)	(V)(SIAHQ)	1.00	0.26	0.23	0.75	14.70
Nenets (3—Aikhenvald 2004: 31)	(V)(SIAHQ)	1.00	0.26	0.23	0.75	14.70

Northern Khanty (3—Aikhenvald 2004: 155-6)	(V)(SIAHQ)	1.00	0.26	0.23	0.75	14.70
Svan (3—Aikhenvald 2004: 76)	(V)(SIAHQ)	1.00	0.26	0.23	0.75	14.70
Turkish (3—Aikhenvald 2004: 155)	(V)(SIAHQ)	1.00	0.26	0.23	0.75	14.70
Xakas (3—Aikhenvald 2004: 30)	(V)(SIAHQ)	1.00	0.26	0.23	0.75	14.70
Meithei (3—Aikhenvald 2004: 31)	(VIAHQ)(S)	1.00	0.27	0.25	0.75	14.70
Siona (3—Aikhenvald 2004: 46)	(V)(S)(IAHQ)	1.29	0.04	0.04	0.95	20.88
Barasano (3—Aikhenvald 2004: 51)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Dulong (3—Aikhenvald 2004: 47)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Eastern Pomo (3—Aikhenvald 2004: 52)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Enets (3—Aikhenvald 2004: 47)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Ladakhi (3—Aikhenvald 2004: 53)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Macuna (3—Aikhenvald 2004: 52)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Maricopa (3—Aikhenvald 2004: 47)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Nganasan (3—Aikhenvald 2004: 47)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Oksapmin (3—Aikhenvald 2004: 46)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Qiang (3—Aikhenvald 2004: 44-45)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Shibacha Lisu (3—Aikhenvald 2004: 54)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Siriano (3—Aikhenvald 2004: 51-2)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Tatuyo (3—Aikhenvald 2004: 51)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Tucano (3—Aikhenvald 2004: 51-3)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Washo (3—Aikhenvald 2004: 46)	(V)(S)(IA)(HQ)	1.35	0.01	0.00	0.97	24.12
Desano (3—Aikhenvald 2004: 60)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Hupa (3—Aikhenvald 2004: 31)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Hupda (3—Aikhenvald 2004: 60)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Kashaya (3—Aikhenvald 2004: 60)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Mamainde (3—Aikhenvald 2004: 56)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Pawnee (3—Aikhenvald 2004: 56)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Southern Nambiquera (3—Aikhenvald 2004: 61)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Tariana (3—Aikhenvald 2004: 1-3;52)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Tsafiki (3—Aikhenvald 2004: 54)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Tuyuca (3—Aikhenvald 2004: 60)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Wichita (3—Aikhenvald 2004: 59-60)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Wintu (3—Aikhenvald 2004: 60)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Xamatauteri (3—Aikhenvald 2004: 18;56)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Yuhup (3—Aikhenvald 2004: 292)	(V)(S)(I)(A)(HQ)	1.37	0.00	0.00	0.98	25.94
Central Pomo (3—Aikhenvald 2004: 61)	(V)(S)(I)(A)(H)(Q)	1.42	0.00	0.01	0.99	32.00
Fasu (3—Aikhenvald 2004: 62-3)	(V)(S)(I)(A)(H)(Q)	1.42	0.00	0.01	0.99	32.00

Table S9. Attested evidentiality systems and their parameters according to the model. V-visual; S-sensory; I-inferred; A-assumed; H-hearsay; Q-quoted.

References

1. GG Corbett, *Number.* (Cambridge University Press), (2000).
2. Ö Dahl, *Tense and Aspect Systems.* (Basil Blackwell), (1985).
3. A Aikhenvald, *Evidentiality.* (Oxford University Press), (2004).
4. R Chen, Ph.D. thesis (Universität Leipzig) (2019).
5. J Brunner, Phonological length of number marking morphemes in the framework of typological markedness in *Between the Regular and the Particular in Speech and Language.* (Peter Lang), pp. 5–28 (2010).
6. M Haspelmath, Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *J. Linguist.* **57**, 605–633 (2021).
7. M Baerman, D Brown, GG Corbett, *Understanding and Measuring Morphological Complexity.* (Oxford University Press), (2015).
8. G Stump, The nature and dimensions of complexity in morphology. *Annu. Rev. Linguist.* **3**, 65–83 (2017).
9. J Nichols, Why is gender so complex? Some typological considerations in *Grammatical gender and linguistic complexity Volume I: General issues and specific studies.* (Language Sciences Press), (2019).
10. GI Bacon, Ph.D. thesis (University of California, Berkeley) (2020).
11. C Bentz, T Soldatova, A Koplenig, T Samardžić, A comparison between morphological complexity measures: Typological data vs. language corpora in *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity.* (2016).
12. C Kemp, T Regier, Kinship categories across languages reflect general communicative principles. *Science* **336**, 1049–1054 (2012).
13. M Denić, S Steinert-Threlkeld, J Szymanik, Complexity/informativeness trade-off in the domain of indefinite pronouns in *Semantics and Linguistic Theory 2020.* (2020).
14. S Steinert-Threlkeld, Quantifiers in natural language optimize the simplicity/informativeness trade-off in *Proceedings of the 22nd Amsterdam Colloquium.* pp. 513–522 (2020).
15. P Grunwald, P Vitányi, Shannon information and Kolmogorov complexity. *arXiv preprint cs/0410002* **1** (2004).
16. N Zaslavsky, C Kemp, T Regier, N Tishby, Efficient compression in color naming and its evolution. *Proc. Natl. Acad. Sci.* **115**, 7937–7942 (2018).

17. G Menardi, N Torelli, Training and assessing classification rules with imbalanced data. *Data Min. Knowl. Discov.* **28**, 92–122 (2014).
18. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria), (2020).
19. M Kuhn, H Wickham, *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.*, (2020).
20. E Hvitfeldt, *themis: Extra Recipes Steps for Dealing with Unbalanced Data*, (2020) R package version 0.1.2.
21. D Bates, M Mächler, B Bolker, S Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
22. N Zaslavsky, T Regier, N Tishby, C Kemp, Semantic categories of artifacts and animals reflect efficient coding in *41st Annual Conference of the Cognitive Science Society.* (2019).
23. JH Greenberg, *Universals of Language.* (MIT press), (1963).
24. JH Greenberg, *Language Universals with Special Reference to Feature Hierarchies.* (The Hague: Mouton), (1966).
25. M Haspelmath, A Karjus, Explaining asymmetries in number marking: Singulatives, pluratives, and usage frequency. *Linguistics* **55**, 1213–1235 (2017).
26. N Zaslavsky, Ph.D. thesis (The Hebrew University of Jerusalem) (2020).
27. B Berlin, P Kay, *Basic Color Terms: Their University and Evolution.* (University of California Press), (1969).
28. N Tishby, F Pereira, W Bialek, The information bottleneck method in *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing.* (1999).
29. JL Bybee, RD Perkins, W Pagliuca, *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World.* (University of Chicago Press), (1994).
30. G Szagun, On the frequency of use of tenses in English and German children's spontaneous speech. *Child Dev.* **49**, 898–901 (1978).
31. W Croft, *Typology and Universals.* (Cambridge University Press), (1990).
32. WA Foley, *The Papuan languages of New Guinea.* (Cambridge University Press), (1986).
33. D Harbour, Ph.D. thesis (Massachusetts Institute of Technology) (2003).
34. D Harbour, Paucity, abundance, and the theory of number. *Language* **90**, 185–229 (2014).
35. B Comrie, *Tense.* (Cambridge University Press), (1985).
36. V Velupillai, Partitioning the timeline: A cross-linguistic survey of tense. *Stud. Lang.* **40**, 93–136 (2016).
37. MK Johnson, S Hashtroudi, DS Lindsay, Source monitoring. *Psychol. Bull.* **114**, 3 (1993).
38. D Saratsli, S Bartell, A Papafragou, Cross-linguistic frequency and the learnability of semantics: Artificial language learning studies of evidentiality. *Cognition* **197**, 104194 (2020).
39. D Gentner, M Bowerman, Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis in *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*, eds. J Guo, et al. pp. 465–480 (2009).
40. WR Miller, *Acoma grammar and texts.* (University of California Press), (1965).
41. K Refsing, *The Ainu Language: The morphology and syntax of the Shizunai dialect.* (Aarhus University Press), (1986).
42. JR Roberts, *Amele.* (Croom Helm), (1987).
43. D Pitman, *Bosquejo de la gramática Araona.* (Instituto Lingüístico de Verano), (1980).
44. RJ Conrad, K Wogiga, *An outline of Bukiyip grammar.* (The Australian National University), (1991).
45. RC Stevenson, *Bagirmi Grammar.* (University of Khartoum), (1969).
46. TG Penchoen, *Tamazight of the Ayt Ndhir.* (Undena Publications), (1973).
47. HC Wolfart, Sketch of Cree, an Algonquian language in *Handbook of North American Indians.* (Smithsonian Institution) Vol. 17, pp. 390–439 (1996).
48. HM Bromley, *A grammar of Lower Grand Valley Dani.* (The Australian National University), (1981).
49. P Newman, *The Hausa Language. An Encyclopedic Reference Grammar.* (Yale University Press), (2000).
50. B Harriehausen, *Hmong Njua.* (Niemeyer), (1990).
51. C Grinevald Craig, *Jacaltec: The Structure of Jacaltec.* (The University of Texas Press), (1977).
52. P Dickens, Jul'hoan orthography in practice. *South Afr. J. Afr. languages* **11**, 99–104 (1991).
53. E Krejnovich, O grammaticheskom vyrazhenii imennykh klassov v glagole ketskogo jazyka in *Studia Ketica: Linguistique.* pp. 139–195 (1968).
54. JM Jacob, *Introduction to Cambodian.* (Oxford University Press), (1968).
55. LJ Watkins, *A grammar of Kiowa.* (University of Nebraska Press), (1984).
56. J Heath, *A grammar of Koyraboro (Koroboro) Senni.* (Köppe), (1999).
57. F Boas, Annotated version of grammar of the Kutenai language, by Pater Philippo Canestrelli; additional notes on the Kutenai language. *Int. J. Am. Linguist.* **45**, 61–94 (1927).
58. E Buechel, *A grammar of Lakota: The language of the Teton Sioux Indians.* (Swift), (1939).
59. MP Noonan, *A Grammar of Lango.* (de Gruyter), (1992).
60. A Terrill, Ph.D. thesis (Australian National University, Canberra) (1999).
61. M Dryer, M Haspelmath, The World Atlas of Language Structures Online (2020).
62. AE Horton, *A grammar of Luvale.* (Witwatersrand University Press), (1949).
63. J Dez, *Structures de la langue malgache.* (Publications orientalistes de France), (1980).

64. CJ Smeets, Ph.D. thesis (Rijksuniversiteit Leiden) (1989).
65. L Gordon, The development of evidentials in Maricopa in *Evidentiality: The linguistic coding of epistemology.* (Abex), pp. 75–88 (1986).
66. PH Dol, Ph.D. thesis (Rijksuniversiteit Leiden) (1999).
67. SL Chelliah, *A Grammar of Meithei.* (de Gruyter), (1997).
68. AHc Chang, Ph.D. thesis (The Australian National University) (2006).
69. DL Everett, Pirahã in *Handbook of Amazonian languages.* (de Gruyter), (1986).
70. C Thornell, Ph.D. thesis (Lund University) (1997).
71. DM Borgman, Sanuma in *The Amazonian Languages*, eds. RMW Dixon, AY Aikhenvald. (Cambridge University Press), (1999).
72. N Kruspe, *A grammar of Semelai.* (Cambridge University Press), (2004).
73. D Wilson, *Suena grammar.* (SIL), (1974).
74. P Schachter, FT Otanes, *Tagalog Reference Grammar.* (University of California Press), (1972).
75. M Donohue, Syntactic categories in Tukang Besi. *Revue québécoise de linguistique* **27**, 71–90 (1999).
76. DH Nguyen, *Vietnamese.* (Benjamins), (1997).
77. D Payne, Yagua in *Handbook ofAmazonian Languages.* (de Gruyter) Vol. 2, (1990).
78. J Bohnemeyer, *The grammar of time reference in Yukatek Maya.* (Lincom), (2002).
79. SdS Facundes, Ph.D. thesis (State University of New York at Buffalo) (2000).
80. J Popjes, J Popjes, Canela-Krahô in *Handbook of Amazonian languages.* (de Gruyter) Vol. 1, pp. 128–199 (1986).
81. DM Topping, BC Dungca, *Chamorro Reference Grammar.* (University of Hawaii Press), (1973).
82. ND Evans, *A grammar of Kayardild.* (de Gruyter), (1995).
83. VZ Panfilov, *Grammatika nivxskogo jazyka.* (1962).
84. DL Everett, B Kern, *Wari: The Pacaas Novos Language of Western Brazil.* (Routledge), (1997).
85. E Maslova, *A grammar of Kolyma Yukaghir.* (de Gruyter), (2003).
86. JC Eisele, *Arabic verbs in time: Tense and aspect in Cairene Arabic.* (Harrassowitz), (1999).
87. TJ Curnow, Ph.D. thesis (The Australian National University) (1997).
88. MS Andronov, *The Brahui Language.* (Nauka), (1980).
89. H Berger, *Die Burushaski-Sprache von Hunza und Nager.* (Harrassowitz), (1998).
90. E Murane, *Daga Grammar.* (SIL), (1974).
91. RMW Dixon, *A grammar of Boumaa Fijian.* (University of Chicago Press), (1988).
92. M Hayashi, B Spreng, Is Inuktitut tenseless? in *Proceedings of the 2005 Annual Meeting of the Canadian Linguistics Association.* (2005).
93. JO Svantesson, Tense, mood and aspect in Mongolian. *Work. Pap. (Lund Univ. Dep. Linguist.* **38**, 189–204 (1991).
94. HS Lee, Ph.D. thesis (University of California, Los Angeles) (1991).
95. M Haspelmath, *A Grammar of Lezgian.* (de Gruyter), (1993).
96. F Merlan, *Mangarayi.* (North Holland), (1982).
97. AC Dench, *Martuthunira: A language of the Pilbara region of Western Australia.* (The Australian National University), (1995).
98. T Salminen, *Tundra Nenets Inflection.* (Suomalais-Ugrilainen Seura), (1997).
99. J Owens, *A Grammar of Harar Oromo (Northeastern Ethiopia).* (Buske), (1985).
100. CG Grinevald, *A grammar of Rama.* (Universite de Lyon), (1990).
101. CR Osborne, *The Tiwi language.* (Australian Institute of Aboriginal Studies), (1974).
102. A Romero-Figueroa, *A reference grammar of Warao.* (Lincom), (1997).
103. JM Dedrick, EH Casad, *Sonora Yaqui Language Structures.* (University of Arizona Press), (1999).
104. R Harrison, M Harrison, C García, *Diccionario zoque de Copainalá.* (SIL), (1981).
105. W Seiler, *Imonda, a Papuan language.* (The Australian National University), (1985).
106. V Du Feu, *Rapanui: A descriptive grammar.* (Routledge), (1996).
107. DC Derbyshire, *Hixkaryana.* (North-Holland), (1979).
108. J Dum-Tragut, *Armenian: Modern Eastern Armenian.* (Benjamins), (2009).
109. MJ Dunn, Ph.D. thesis (1999).
110. LD Lohitare, DL Lohammarimoi, DT Peter, PL Joseph, *Didinga Grammar Book.* (2012).
111. W McGregor, *A Functional Grammar of Gooniyandi.* (Benjamins), (1990).
112. P Svenonius, Icelandic case and the structure of events. *The J. Comp. Ger. Linguist.* **5**, 197–225 (2002).
113. SN Sridhar, *Kannada.* (Routledge), (1990).
114. J Hewson, V Bubenik, *Tense and Aspect in Indo-European languages: Theory, typology, diachrony.* (Benjamins), (1997).
115. M Davidson, Ph.D. thesis (State University of New York at Buffalo) (2002).
116. MA Macaulay, *A Grammar of Chalcatongo Mixtec.* (University of California Press), (1996).
117. L Drapeau, *Grammaire de la langue innue.* (Presses de l'Université du Québec), (2014).
118. T Donaldson, *Ngiyambaa: The language of the Wangaaybuwan.* (Cambridge University Press), (1980).
119. J Saeed, *Somali.* (Benjamins), (1999).
120. A Lindfors, Ph.D. thesis (Uppsala University) (2003).

121. DS Rood, *Wichita grammar.* (Garland), (1976).
122. M Lamberti, R Sottile, *The Wolaytta Language.* (Köppe), (1997).
123. JD Sapir, *A Grammar of Diola Fogny.* (Cambridge University Press), (1965).
124. KJ Franklin, *A grammar of Kewa, New Guinea.* (The Australian National University), (1971).
125. I Nedjalkov, *Evenki.* (Routledge), (1997).
126. RS Hagman, *Nama Hottentot Grammar.* (Indiana University), (1977).
127. C Voorhoeve, *The Flamingo Bay dialect of the Asmat language.* (The Hague), (1965).
128. L Bruce, *The Alamblak language of Papua New Guinea (East Sepik).* (The Australian National University), (1984).
129. W Jones, P Jones, *Barasano syntax.* (SIL), (1991).
130. GD Kimball, Ph.D. thesis (Tulane University) (1985).
131. D Wdzenczny, Temporal expression in Wichí nominals. *Santa Barbar. Pap. Linguist.* **22** (year?).
132. G Innes, *An introduction to Grebo.* (University of London (School of Oriental & African Studies)), (1966).
133. R Carlson, *A grammar of Supyire.* (de Gruyter), (1994).
134. N Akiyama, C Akiyama, *Japanese grammar.* (Simon & Schuster), (2012).
135. F Boas, EC Deloria, *Dakota grammar.* (US Government Printing Office) Vol. 23, (1941).
136. RD Bugenhagen, *A grammar of Mangap-Mbula: An Austronesian language of Papua New Guinea.* (The Australian National University), (1991).
137. W Chafe, *A grammar of the Seneca language.* (University of California Press), (2015).
138. P Cole, *Imbabura Quechua.* (North Holland), (1982).
139. N Fallou, *Wolof.* (Lincom), (2003).
140. M Fortescue, *West Greenlandic.* (Croom Helm), (1984).
141. L Grout, *The Isizulu: a grammar of the Zulu language.* (Trübner & Company), (1859).
142. MJ Hardman, Data-source marking in the Jaqi languages in *Evidentiality: The linguistic coding of epistemology*, eds. WL Chafe, J Nichols. (1986).
143. L Jeon, J Li, S Mauney, A Navarro, J Wittke, A basic sketch grammar of Gĩkũyũ. *Rice Work. Pap. Linguist.* **6** (2015).
144. GD Kimball, *Koasati grammar.* (University of Nebraska Press), (1991).
145. A Kunnap, On the Enets evidential suffixes. *Linguist. Uralica* **38**, 145–154 (2002).
146. DP Lombard, *Introduction to the grammar of Northern Sotho.* (van Schaik), (1985).
147. I Nikolaeva, *A grammar of Tundra Nenets.* (De Gruyter Mouton), (2014).
148. DR Parks, *A grammar of Pawnee.* (Garland), (1976).
149. D Praulinš, *Latvian: An essential grammar.* (Routledge), (2012).
150. C Rounds, *Hungarian: An essential grammar.* (Routledge), (2009).
151. HF Schiffman, F Harold, *A reference grammar of spoken Tamil.* (Cambridge University Press), (1999).
152. J Swanton, *Haida.* (US Government Printing Office), (1911).
153. R Underhill, *Turkish grammar.* (MIT Press) Vol. 460, (1976).
154. V Velupillai, *Hawai'i Creole English: A typological analysis of the tense-mood-aspect system.* (Springer), (2016).
155. J Yeon, L Brown, *Korean: A comprehensive grammar.* (Routledge), (2013).