Grammatical marking and the tradeoff between code length and informativeness

Francis Mollica (mollicaf@gmail.com)[1]
Geoff Bacon (bacon@berkeley.edu)[2]
Yang Xu (yangxu@cs.toronto.edu)[3]
Terry Regier (terry.regier@berkeley.edu)[2,4]
Charles Kemp (c.kemp@unimelb.edu.au)[1]

[1] School of Psychological Sciences, University of Melbourne, Parkville, Victoria 3010, Australia
[2] Department of Linguistics, University of California, Berkeley, CA 94720 USA
[3] Department of Computer Science, Cognitive Science Program, University of Toronto, Toronto, ON M5S 3G8 Canada
[4] Cognitive Science Program, University of California, Berkeley, CA 94720 USA

Abstract

Functionalist accounts of language suggest that forms are paired with meanings in ways that support efficient communication. Previous work on grammatical marking suggests that word forms have lengths that enable efficient production, and previous work on the semantic typology of the lexicon suggests that word meanings represent efficient partitions of semantic space. Here we consider an integrated information-theoretic framework that captures how communicative pressures influence both form and meaning. We take tense systems as a case study, and show how the framework explains both which tense systems are attested across languages and the length asymmetries of the forms in those systems.

Keywords: efficient communication; semantic typology; grammar; information theory

Functionalist linguistics has a long history of exploring ways in which languages support efficient communication, and this perspective has been applied to many areas including phonology, morphology, syntax, and semantics (e.g. Croft, 2003). Computational work in this tradition often characterizes communicative efficiency using information theory (e.g. Gibson et al., 2019), and this approach is appealing in part because information theory captures general computational principles that potentially apply across many levels of linguistic structure. Here we show that an existing information-theoretic account of lexical semantics (Kemp, Xu, & Regier, 2018, as formulated by Zaslavsky, Kemp, Regier, & Tishby, 2018) also accounts for classic ideas about coding efficiency from the literature on grammatical marking (Greenberg, 1966; Haspelmath, 2018). Connecting these lines of work illustrates how information theory provides a unified account of both the meanings encoded in natural language and the forms used to express them.

The information-theoretic framework that we use formalizes the tradeoff between informativeness and simplicity that languages must negotiate. Consider a speaker who wishes to convey some meaning (e.g. the temporal location of an event) to a hearer. A highly informative system allows the speaker to discriminate between many different meanings (e.g. many temporal intervals), but this communicative precision can only be achieved if the system is far from simple. The tradeoff between informativeness and simplicity has been discussed for many years in the literature on "competing motivations" (Haiman, 2010) and here we build on a recent account of color-naming across languages that formalizes both dimensions in information-theoretic terms (Zaslavsky et al., 2018).

We show how the simplicity dimension connects naturally with the notion of coding efficiency from the literature on grammatical marking (Greenberg, 1966; Hawkins, 2004). Haspelmath (2018) reviews a substantial body of evidence suggesting that grammatical constructions across languages are well-adapted in the sense that they tend to minimize expected code length. Haspelmath (2018) focuses primarily on the *forms* used to encode number, case, tense, and other grammatical domains, but we will suggest that the tradeoff between informativeness and simplicity helps to explain both which grammatical categories are attested across languages and the relative lengths of the forms used for these categories.

Although the tradeoff between informativeness and simplicity has implications for many aspects of grammar, we focus on tense as a case study. Languages have many different ways of locating events in time, and we focus specifically on grammatical categories that play this role. English, for example, has grammatical categories that distinguish between past, present and future. The expressions "she walked," "she walks" and "she will walk" distinguish between past, present and future using inflectional morphology ("walk" + "-ed", "walk" + "-s") and periphrasis ("will" + "walk"). Some languages, however, have more elaborate tense systems that specify not only whether an event is in the past or future, but also how far in the past or future it is. For example, Hixkaryana distinguishes between events in the immediate past (same day or previous night), near past (past few months) and remote past (Derbyshire, 1979).

The next section introduces the theoretical framework we use and provides formal definitions of informativeness and complexity (the inverse of simplicity). We then apply the framework to tense, and show how tense systems across many languages achieve near-optimal tradeoffs between informativeness and simplicity. The final sections of the paper show how our approach connects with existing work on coding asymmetries in grammatical marking, and discuss the benefits gained by integrating functionalist approaches to form and meaning.

Theoretical framework

We adopt the theoretical framework of Zaslavsky et al. (2018), who explored the efficiency of semantic systems in the specific case of color naming. We show that the same framework can also be linked to aspects of linguistic form. The framework assumes a speaker and a listener who wish to communicate about objects in a conceptual universe U. The speaker has a particular meaning $m \in M$ in mind, which takes the form of a distribution over objects in U. To express this meaning, the speaker produces a word or other form $w \in W$ according to an encoder $q(w|m)$, which stochastically maps meanings m into forms w. Upon hearing the speaker utter w, the listener computes a distribution $\hat{m}_w$ that is intended to approximate the speaker's meaning m, via Bayesian inference using the encoder $q(w|m)$ and a prior distribution $p(m)$ over meanings. Let the encoder $q(w|m)$ and the prior $p(m)$ together constitute a grammar $\mathcal{G}$ that supports communication between the speaker and listener. The efficiency of such a grammar is then determined by two quantities: informativeness and complexity.

Informativeness is inversely related to the *communicative cost* $C_C(\mathcal{G})$ of the grammar. Following Regier, Kemp, and Kay (2015) and Zaslavsky et al. (2018) we define this cost as the expected Kullback-Leibler (KL) divergence $D(m||\hat{m}_w)$ between the mental representation m of the speaker and the listener's mental reconstruction $\hat{m}_w$ of that meaning:

$$C_C(\mathcal{G}) = E[D(m||\hat{m}_w)] = \sum_{m,w} p(m)q(w|m)D(m||\hat{m}_w). \quad (1)$$

Again following Zaslavsky et al. (2018), the complexity $C_L(\cdot)$ of a grammar is defined as the mutual information between forms and meanings:

$$C_L(\mathcal{G}) = I(M;W) = H(W) - H(W|M), \quad (2)$$

which, as shown, can be formulated as the difference of two terms: the entropy over words $H(W)$ and the conditional entropy of words given meanings $H(W|M)$. This formulation sets the stage for the link to linguistic form. The entropy over words $H(W)$ captures the expected amount of information needed to represent a word and, thus, the expected code length assuming an optimal code that uses $-\log(P(w))$ bits to represent an event with probability $P(w)$. If the mapping between words and meanings is deterministic (as is often the case in grammatical systems), then $H(W|M)$ is zero and our complexity measure reduces to $H(W)$, which has the link to code length stated above.

All grammatical systems must negotiate the tradeoff between complexity and informativeness, and the information bottleneck method (Tishby, Pereira, & Bialek, 1999; Zaslavsky et al., 2018) allows us to characterize the optimal tradeoff between these dimensions. In order to apply the method we must specify the conceptual universe of possible time intervals U, the meaning distributions $p(u|m)$, and the probability $p(m)$ of needing to communicate each meaning.

Tense

The conceptual universe. Work in formal semantics has produced precise representations for tense that could potentially be used in frameworks like ours (e.g., Reichenbach, 1947). As a starting point, however, we considered a simple discretized timeline, and focused on absolute tense: the grammatical expression of time relative to the present.[1] We defined the conceptual universe U as a discrete timeline with seven temporal intervals: remote past (a), near past (b) and immediate past (c), present (r), immediate future (x), near future (y) and remote future (z). These intervals are not sufficient to capture the tense system of every language in full: for example, Comrie (1985) reports that Kiksht, a language of the US Pacific Northwest, distinguishes between six or seven past tense categories. Our seven-interval timeline is therefore a pragmatic choice that allows us to represent the tense systems of many but not all of the languages of the world.

We compiled a data set of tense systems for 159 languages. 53 systems were drawn from Dahl (1985) and the rest from a range of other resources.[2] For each language, we coded the system's categories on our discrete timeline and noted if any categories in the system were zero-marked: that is, expressed by default without requiring any additional grammatical markers. For example, in Afrikaans the present ("Ek loop", meaning "I walk") can be expressed without adding any markers to the base form of the verb, but the past ("Ek het geloop") and future ("Ek sal loop") both require additional markers. In some languages a tense is explicitly marked only for some combinations of person and number: for example, in English the present requires the marker "-s" in "she walks" but not in "I walk." We treat a tense as zero-marked only if it is unmarked for all combinations of person and number.

All systems that occur at least twice in the data are listed in Table 1. The representations in the first column use a, b and c to denote the three past intervals, r to denote the present, and x, y, and z to denote the three future intervals. Each explicitly marked category within the system is enclosed in parentheses. For example, (abc)(r)(xyz) is the system that includes marked categories for past, present and future. In our notation the absence of parentheses denotes zero marking. For example, (abc)(r)(xyz) is used for languages like English that explicitly mark past, present and future, and (abc)r(xyz) is used for languages like Afrikaans that include the same three categories but zero-mark the present. Our coding assumes that all cultures share the need to communicate and mentally represent all seven of the intervals in U. As a result, languages without tense marking are encoded as a system abcrxyz with a zero-marked category that includes every interval in U.

A major challenge encountered in compiling the data

---

[1]We leave as an important open question whether the ideas in this paper will generalize to relative tense.

[2]Source list and data can be found at `https://osf.io/jgdp4/`

system	count	communicative cost	complexity	expected length	frontier dist	instance
abcrxyz	47	1.0142	0.0000	0.000	0.0000	Cebuano
(abc)(r)(xyz)	39	0.0846	1.5225	1.000	0.0001	English
(abc)rxyz	20	0.4723	0.8471	0.274	0.0421	Burushaski
abcr(xyz)	13	0.4968	0.0444	0.251	0.0444	Indonesian
(abc)r(xyz)	11	0.0846	1.5225	0.525	0.0001	Maltese
(abc)(rxyz)	6	0.4723	0.8471	1.000	0.0421	Brahui
(abcr)(xyz)	5	0.4968	0.8129	1.000	0.0444	Seneca
(a)(b)(c)r(xyz)	3	0.0373	1.9314	0.525	0.0124	Koasati
(ab)(c)r(x)(yz)	2	0.0198	2.0474	0.525	0.0047	Supyire
(ab)(c)(r)(xyz)	2	0.0477	1.7965	1.000	0.0089	Sesotho
(a)(bc)r(xyz)	1	0.0505	1.7576	0.525	0.0066	Khoekhoe
(a)(b)(c)(rxyz)	1	0.4250	1.2561	1.000	0.2078	Hixkaryana

Table 1: All tense systems that occur two or more times in our data, along with two that appear once (Khoekhoe and Hixkaryana) and are mentioned in the text. The notation for tense systems is described in the text, and distinguishes between systems that include the same categories but are different with respect to zero marking. The frontier dist column shows the distance between each system and the optimal frontier in Figure 3a, and expected length corresponds to the y-axis of Figure 3b.
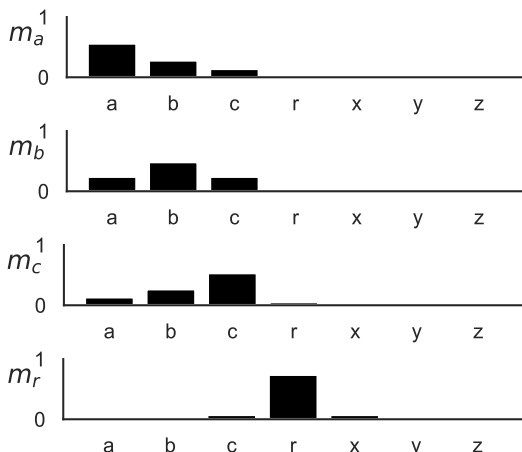

Figure 1: Meanings for model with the 7 element timeline. Meanings when communicating about the future (m_x, m_y, m_z) are not shown here but are mirror images of meanings for communicating about the past (m_c, m_b and m_a).

is that tense is often hard to separate from aspect and modality (Bybee, Perkins, & Pagliuca, 1994). For example, in some languages the primary distinction is between perfective and imperfective (roughly whether an action is complete or incomplete) rather than between past and future. Some languages include markers for categories (e.g. past perfective) that combine tense and aspect. When consulting our primary sources, we made our best judgment about whether a language could be represented in our coding scheme without distorting it too greatly, and excluded a small number of languages (e.g. Hawaiian) for which our scheme seemed especially inadequate.

A second and less fundamental challenge is that languages which make remoteness distinctions do not include categories that are precisely equivalent. Our coding scheme distinguishes remote (more than 7 days distant), near (between 1 and 7 days) and immediate (on the same day), and we fitted each language into this scheme as best we could.

Figure 2: Prior distribution on the seven intervals along the timeline.

Domain structure. Our framework formalizes meanings as probability distributions over intervals in the time line. These distributions can capture uncertainty about the interval to which an event belongs. For example, a speaker might not be sure whether a given state of affairs has ceased or is ongoing. This uncertainty should respect the underlying structure of the domain: for example, a speaker should be more likely to confuse present with past than future with past. We therefore use a hierarchy that postulates major boundaries between past, present and future, and minor boundaries between the three pasts (remote, near and immediate) and between the three futures. Figure 1 shows meanings defined over the seven-element timeline. The distributions are defined in terms of two parameters λ and μ that specify how sharply probability mass decreases across minor and major boundaries. We set $\lambda = 0.5$ and $\mu = 0.1$, which means that distributions drop by factors of 2 and 10 across minor and major boundaries, respectively.

Communicative need. As described above, our formal framework requires a probability distribution $p(m)$ that captures how often speakers attempt to convey each of the seven different meanings. Given the lack of large-scale, annotated corpora, we estimated these probabilities using a two-step process. In the first step we used estimates of past, present and future from an analysis of social media (Park et al., 2017). The resulting counts yield a distribution of $[0.274, 0.475, 0.251]$ over the coarse categories of past, present and future. There is a clear preference for present over either past or future, and also a weaker preference for past over future.
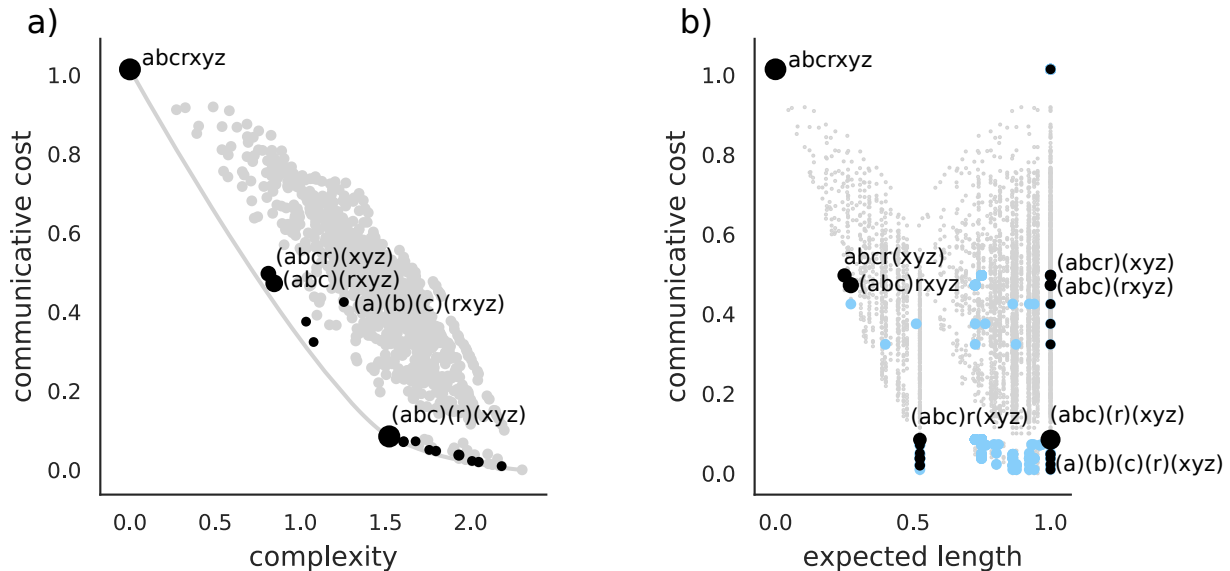
Figure 3: Tradeoff plots for the model. (a) Black dots represent attested systems (size denotes frequency) and grey dots represent all possible deterministic systems. (b) Light blue dots represent all ways to zero-mark at most one category in an attested system, and small grey dots show all ways to apply zero-marking to unattested systems. The column of attested systems with expected length equal to 1 includes systems that do not use zero marking.

Deg. of remoteness	Temporal adverbs
immediate (c, x)	*today*
near past (b)	*yesterday*
near future (y)	*tomorrow*
remote past (a)	*last week/month/year/decade/century*
remote future (z)	*next week/month/year/decade/century*

Table 2: Temporal adverbs used to estimate usage probabilities for intervals along the timeline.

Second, we used frequencies of the temporal adverbs shown in Table 2 to distribute probability mass among the three levels of remoteness within both past and future categories. All frequencies were derived from the Google N-gram English corpus (Michel et al., 2011) for 1985. 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 intervals. The prior distribution resulting from the entire two-step process is shown in Figure 2.

Analyses and results

Figure 3a shows the tradeoff plot for the model. Attested systems are plotted in black with sizes reflecting the frequency of each system in our language sample. For comparison, the light grey points show all possible, yet unattested, deterministic tense systems in our coding. The grey curve shows the optimal trade-off between communicative cost and complexity. The figure reveals that all ten systems in Table 1 that occur two or more times in the data set lie close to the optimal frontier. One attested system (Hixkaryana in Table 1) lies further from the optimal

frontier than all of the others. This system is unusual because it includes a relatively large number of categories but does not distinguish between present and future.

Figure 3a is based on category extensions rather than the forms for these categories, but we also wished to explore the link to linguistic form mentioned above. The complexity measure on the x-axis of Figure 3a corresponds to expected code length only if the systems are deterministic and if $-\log(P(w))$ bits are used to represent a word with probability $P(w)$. Piantadosi, Tily, and Gibson (2011) show that word lengths and predictability are correlated, and in principle one could test the prediction that the lengths of tense markers are proportional to their prior probabilities. Our dataset does not include the forms of tense markers but is annotated for zero-marking, which allows us to conduct a much coarser analysis. For any given system, suppose that at most one of the categories can be zero marked and that forms of unit length are used for all remaining categories. Figure 3b shows the trade-off between informativeness and expected form length under this assumption. The black dots represent attested systems, and the light blue dots include all systems that use zero marking for at most one category in an attested system. The small grey dots show all ways to apply zero-marking to unattested systems. A number of languages do not use zero-marking at all, and these languages appear as a column of black dots with expected length equal to one. Attested systems that do use zero-marking all occur towards the left of the plot, revealing that systems with zero marking overwhelmingly tend to zero-mark the most frequent category.

Although the representations in our data set specify hard category boundaries (i.e. are deterministic), in reality these
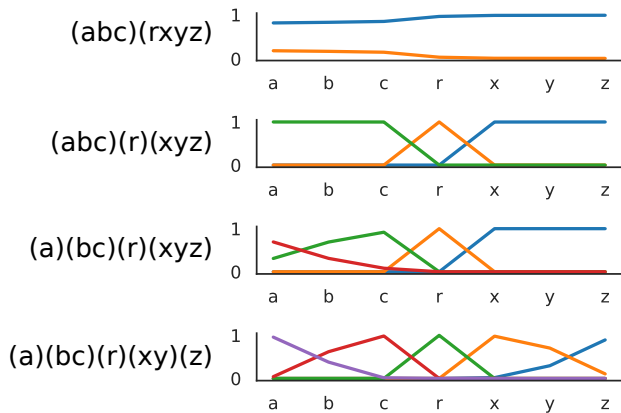
Figure 4: Evolutionary sequence generated by traversing the optimal frontier in Figure 3a. Each panel shows curves $P(c|e)$ for each category c in a system, where e ranges over the 7 intervals in the timeline. The system labels at the left are based on the deterministic systems that best correspond to the 4 stochastic systems plotted here.

boundaries are often soft and probabilistic. For example, if a language marks near and remote past differently, an event one week ago might sometimes be treated as near and sometimes treated as remote. Most of the systems along the optimal frontier in Figure 3a have soft category boundaries, and Figure 4 shows a sequence of systems that emerge as the frontier is traversed from top left to bottom right, by analogy with the trajectory of optimal color naming systems obtained in the same manner by Zaslavsky et al. (2018). The first pattern to emerge is a soft distinction between past and non-past, and a more discrete version of this two-category system matches the most frequent two-category system in our data. Next present is separated from future, and the resulting system is by far the most frequent three-category system in our data. The next distinction to emerge separates remote and non-remote pasts. The resulting system is similar to Khoekhoe in Table 1, but our data are insufficient to tell whether remote past is attested more often than immediate past. The final system in Figure 4 splits both past and future into near and remote categories. Similar five-category systems are attested in the literature (Sarvasy, 2017; Savić, 2017), but there are none in our data set.

The theoretically optimal systems in Figure 4 are consistent with two previously-identified generalizations about tense systems. Comrie (1985, p 50) suggests that discontinuous tense categories are rare or unattested, and the categories in Figure 1 all correspond to connected regions of the timeline. This property of the model is a consequence of the fact that the meanings in Figure 1 all vary smoothly over the timeline. Comrie (1985, p 85) also points out the "general tendency of languages to have a better developed past than future system," and in Figure 4 the past is both picked out and subdivided before the future. This property of the model follows from the empirically obtained asymmetric prior in Figure 2, which suggests that speakers tend to refer more often to the past than to the future.

Binary Coding Asymmetries: Present vs Future

The zero-marking analysis in Figure 3b suggests how our framework makes predictions about length asymmetries in grammatical forms. These asymmetries have been discussed by Haspelmath (2018) and others, but previous work often focuses on a single binary distinction (e.g. present vs future) at a time. We now show how a simplified version of our approach can be used to analyze binary distinctions, and use this analysis to explain why meaning becomes important when accounting for grammatical systems that make more than two distinctions.

Haspelmath (2018) presents 25 cases in which frequencies appear to predict code lengths (i.e. the lengths of linguistic forms), and one of these cases concerns present and future tenses. This particular example follows Greenberg (1966), who noted that present tense forms tend to be more frequent than future tense forms, and also tend to be shorter. In our terms, suppose that the conceptual universe U is reduced to a set including just two temporal intervals: r (present) and f (future). There are two meanings m_r and m_f that correspond to these two intervals, and we assume for now that these meanings reflect certainty on the part of the speaker (i.e. m_r is a distribution that assigns probability only to the interval r). Based on the prior in Figure 2, we assume that $P(m_r) = 0.65$ and $P(m_f) = 0.35$.

We begin by considering a simplified case in which there is no zero-marking. If the speaker is deterministic, then there are two possible category systems. Let (rf) denote the system that uses the same label for both meanings, and (r)(f) denote the system that uses different labels for the two. Figure 5a shows communicative costs and complexities for the two systems, and the size of each black dot indicates the number of languages in our dataset that lie at that point when considering only marking for present and future. Neither system dominates the other: (r)(f) is better with respect to communicative cost, and (rf) is better with respect to complexity.

To capture Haspelmath (2018)'s coding asymmetries, we consider the zero-marking analysis and the resulting tradeoff plot in Figure 5b. Of the three systems that distinguish between present and future, r(f) (i.e., the system that zero-marks the present) is superior to the unattested system (r)f (light blue dot) and (r)(f) (a system that does not use zero marking).

The formal analyses summarized by Figures 5a and 5b add little to the informal accounts of present and future tenses previously given by Greenberg (1966) and Haspelmath (2018). Critically, however, the analyses can be scaled up to cases where the conceptual universe U includes more than two temporal intervals, and where there is an ordering over these intervals.

Beyond Binary Distinctions

Suppose now that the conceptual universe $U = \{s, r, f\}$ includes past (s) in addition to present and future. As
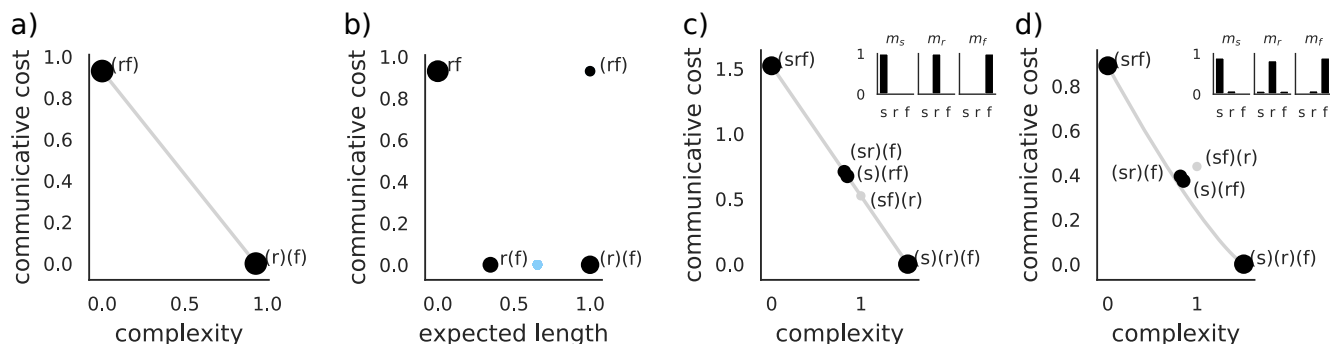
Figure 5: (a) Tradeoff between communicative cost and complexity for a timeline that includes present and future only. (b) Tradeoff between communicative cost and expected code length when zero-marking is permitted. The light blue dot represents the unattested system (r)f. (c) Results for a three element timeline with no domain structure. The inset shows deterministic meanings m_s, m_r and m_f. (d) Results for a three element timeline with domain structure captured by graded meanings (inset). In (a), (c-d) the grey line shows optimal tradeoffs between the two dimensions that can be achieved by stochastic systems.

before, we use three meanings m_s, m_r and m_f that reflect speaker certainty, and set the prior distribution $P(m)$ to $[0.274, 0.475, 0.251]$ based on Figure 2.

Figure 5c shows the resulting tradeoff plot. In Figure 5c, there are five possible systems that partition past, present and future into categories, and all of them lie along the optimal frontier (i.e. none dominates any of the others). Four of these systems are attested across languages, but the "unnatural" system (sf)(r) that groups past and future into a single category that excludes the present is extremely rare or unattested. The model does not rule this system out because it has no ordering over temporal intervals—in particular, it does not know that past is "closer" to present than it is to future. Following our large-scale analysis we can incorporate the structure of the domain using the graded meanings in the inset of Figure 5d ($\mu = 0.1$ as before). The tradeoff plot in Figure 5d now shows that the unnatural system (sf)(r) is dominated by the system (s)(rf), and that all four systems near the optimal frontier are attested.

Discussion and Conclusion

Our results suggest that tense systems across languages achieve efficient tradeoffs between informativeness and complexity, and align with similar results previously reported for domains including color naming (Zaslavsky et al., 2018), kin naming (Kemp & Regier, 2012) and systems of quantifiers (Steinert-Threlkeld, 2020). Our work also demonstrates that existing work on lexical typology (Kemp et al., 2018) and grammatical marking (Haspelmath, 2018) can be usefully brought together. Doing so adds something to both lines of work. Among previous information-theoretic work on the lexicon there are studies that focus on word meaning (e.g. Kemp et al., 2018) and studies that focus on word forms (e.g. Piantadosi et al., 2011), but few that address both meaning and form. Our work suggests how form and meaning can be brought together in an integrated information-theoretic framework.

Previous work on asymmetries in grammatical marking

focuses mainly on the forms used to encode grammatical categories, and highlights the principle that these forms serve to minimize expected code length. Our work suggests that this principle should not be considered in isolation, but instead trades off against other principles such as a pressure for semantically informative communication. Haspelmath (2018) does appeal to the notion of competing motivations and suggests that complexity trades off against *explicitness*, which is the general preference "to express grammatical meanings explicitly." Explicitness may seem conceptually related to semantic informativeness but the two are distinct. Explicitness is purely about form, and is needed to explain why some languages do not make use of zero marking. Semantic informativeness is purely about meaning, and captures the need for meanings to support precise communication.

A possible reason why informativeness is not more prominent in existing previous treatments of grammatical marking is that they often focus on binary oppositions (e.g. present vs future, or present vs past). In these cases, any coding system that marks the distinction in any way is maximally informative, meaning that complexity (i.e. minimizing code length) is the main issue of interest. The tradeoff between complexity and informativeness becomes especially clear when developing an integrated account of a domain (e.g. tense) with more than two elements. In this case a successful theory must explain both which systems are frequently found across languages as well as the relative lengths of forms for the categories in these systems.

A natural next step is to move beyond tense and consider additional grammatical domains. Evidentiality, number, and person are especially interesting because all of these domains have been characterized using scales or hierarchies with more than two elements. Previous work on grammatical marking has demonstrated that the principle of minimizing code length operates across many grammatical domains (Haspelmath, 2018), and we are optimistic that our approach will prove to be comparably broad in scope.

References

Bybee, J. L., Perkins, R. D., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world* (Vol. 196). University of Chicago Press.

Comrie, B. (1985). *Tense*. Cambridge University Press.

Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge University Press.

Dahl, Ö. (1985). Tense and aspect systems.

Derbyshire, D. C. (1979). *Hixkaryana* (Vol. 1). North-Holland.

Gibson, E., Futrell, R., Piantadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy., R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*.

Greenberg, J. H. (1966). *Language universals*. The Hague: Mouton de Gruyter.

Haiman, J. (2010). Competing motivations. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology.*

Haspelmath, M. (2018). Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability.

Hawkins, J. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*, 1049–1054.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*, 109-128.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., . . . others (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Park, G., Schwartz, H. A., Sap, M., Kern, M. L., Weingarten, E., Eichstaedt, J. C., . . . others (2017). Living in the past, present, and future: Measuring temporal orientation with language. *Journal of Personality*, *85*(2), 270–280.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.

Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. *The handbook of language emergence*, *87*, 237.

Reichenbach, H. (1947). The tenses of verbs. In *The language of time: a reader* (pp. 1–12).

Sarvasy, H. (2017). *A grammar of Nungon: A Papuan language of northeast New Guinea*. Brill.

Savić, S. (2017). The perfective and imperfective aspects in Xhosa. *Stellenbosch Papers in Linguistics Plus*, *52*(1), 45–72.

Steinert-Threlkeld, S. (2020). Quantifiers in natural language optimize the simplicity/informativeness trade-off. In *Proceedings of the 22nd Amsterdam colloquium* (p. 513-522).

Tishby, N., Pereira, F., & Bialek, W. (1999). The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing.*

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*.