MORPHOLOGICAL ORGANIZATION: THE LOW CONDITIONAL ENTROPY CONJECTURE

Farrell Ackerman
University of California, San Diego

Robert Malouf
San Diego State University

Crosslinguistically, inflectional morphology exhibits a spectacular range of complexity in both the structure of individual words and the organization of systems that words participate in. We distinguish two dimensions in the analysis of morphological complexity. Enumerative complexity (E-complexity) reflects the number of morphosyntactic distinctions that languages make and the strategies employed to encode them, concerning either the internal composition of words or the arrangement of classes of words into inflection classes. This, we argue, is constrained by integrative complexity (I-complexity). The I-complexity of an inflectional system reflects the difficulty that a paradigmatic system poses for language users (rather than lexicographers) in information-theoretic terms. This becomes clear by distinguishing average paradigm entropy from average conditional entropy. The average entropy of a paradigm is the uncertainty in guessing the realization for a particular cell of the paradigm of a particular lexeme (given knowledge of the possible exponents). This gives one a measure of the complexity of a morphological system—systems with more exponents and more inflection classes will in general have higher average paradigm entropy—but it presupposes a problem that adult native speakers will never encounter. In order to know that a lexeme exists, the speaker must have heard at least one word form, so in the worst case a speaker will be faced with predicting a word form based on knowledge of one other word form of that lexeme. Thus, a better measure of morphological complexity is the average conditional entropy, the average uncertainty in guessing the realization of one randomly selected cell in the paradigm of a lexeme given the realization of one other randomly selected cell. This is the I-complexity of paradigm organization. Viewed from this information-theoretic perspective, languages that appear to differ greatly in their E-complexity—the number of exponents, inflectional classes, and principal parts—can actually be quite similar in terms of the challenge they pose for a language user who already knows how the system works. We adduce evidence for this hypothesis from three sources: a comparison between languages of varying degrees of E-complexity, a case study from the particularly challenging conjugational system of Chiquihuitlán Mazatec, and a Monte Carlo simulation modeling the encoding of morphosyntactic properties into formal expressions. The results of these analyses provide evidence for the crucial status of words and paradigms for understanding morphological organization.*

Keywords: inflectional paradigms, information-theoretic measures, word-based morphology, morphological typology, word-and-paradigm models, morphological complexity

1. Introduction. Theoretical analysis begins with the identification of a question. The question we address here is what Ackerman and colleagues (2009) called the paradigm cell filling problem: what licenses reliable inferences about the surface word forms for the inflectional (and derivational) families of word forms associated with (classes of) lexemes? That speakers are able to do this is a truly puzzling accomplishment given the extraordinary variation and complexity attested in the morphological systems of the world. But the task is, of course, aided in each specific language by the distinctive organization of its inflectional system.[1]

---

* The authors would like to thank Adam Albright, Harald Baayen, Jim Blevins, Olivier Bonami, Jeremy Boyd, Giles Boyé, Raphael Finkel, Alice Harris, Scott Seyfarth, Andrea Sims, Greg Stump, and Greg Carlson and two anonymous referees for their comments and criticisms. Earlier versions of this work were presented at the Workshop on Morphology and Formal Grammar, Paris (2010); the annual meeting of the Linguistic Society of America, Baltimore (2010); the Workshop on Morphological Complexity: Implications for the Theory of Language, Harvard (2010); the Workshop on Quantitative Measures in Morphology and Morphological Development, UCSD (2011); and the Workshop on Challenges of Complex Morphology to Morphological Theory, Boulder (2011). We are grateful to the audiences for their feedback.

[1] That neither the problem nor the basic insight concerning its potential solution in terms of systemic morphological organization is completely novel is evident from the fact they both appear in such a classic work as Paul 1891 as well as in other early research reviewed in Esper 1973.

Languages with inflectional morphology, however complex, exhibit patterns of combination responsible for the internal composition of words as well as for ways in which the members of (classes of) words exhibit relations that cohere into systems of inflectional organization. On its face this observation claims nothing substantive about either the specific nature of the internal organization of words or the specific nature of the system of inflectional organization. Specific hypotheses about them can serve as the bases for the construction of different types of theories.

A familiar and widespread approach to morphology, adapting the tradition of American structuralists and certain ideas of Europeans such as Badouin de Courtenay (Anderson 1985, Hockett 1987, Matthews 2001, among others), focuses on the decomposition of complex words into small meaningful pieces, that is, morphemes. This approach can be characterized as SYNTAGMATIC, because it emphasizes the linear combination of constitutive elements, and COMPOSITIONAL, because it endeavors to derive the meaning of the whole word from the meanings that inhere in its identifiable parts. This morpheme-based conception of the enterprise naturally leads to certain research questions, while precluding others. In particular, it leads to familiar efforts to segment words into stems and morphosyntactic markers as well as the rules (morphotactic and phonological) that yield the legal combinations of these entities in their surface realizations as word forms. It has also led to questions about parsimony regarding the minimal elements (either underlying or surface-based) and operations required to construct or build word forms. From this perspective neither surface word forms nor the systematic patterns of relations that whole words display are construed as basic units of grammatical organization. Rather, the surface words of particular languages are useful to the degree that they provide insight into underlyingly invariant atoms of analysis and combinatoric operations that can account for (variable) surface expression. This reflects a view whereby surface patterns of both words and networks of words (i.e. relations between surface alternants) are not regarded as proper objects of linguistic analysis, while the abstract elements and operations responsible for constructing these ephemera are. This, in turn, naturally leads to questions about the psychological or biological basis of these constitutive elements and operations.

Chomsky (1965:174) expressed suspicion about the contemporaneous fashion for syntagmatic approaches to morphology and a prescient preference for the paradigmatic nature of morphology:

> I know of no compensating advantage for the modern descriptive reanalysis of traditional paradigmatic formulations in terms of morpheme sequences. This [morphemic analysis—*FA/RM*] seems, therefore, to be an ill-advised theoretical innovation … It seems that in inflectional systems, the paradigmatic analysis has many advantages and is to be preferred … It is difficult to say anything more definite, since there have been so few attempts to give precise and principled description of inflectional systems in a way that would have some bearing on the theoretical issues involved here.

He is making reference here to another morphological tradition—one that is often less familiar, though its historical antecedents are long and rich—that focuses on words and the ways in which related surface word forms cohere into networks. Words, in this view, are fundamental, empirical objects of morphological analysis.

This approach can be characterized as PARADIGMATIC, because it identifies (sets of) patterns that whole words participate in, and CONFIGURATIVE, because, while the meaning of a word form is not necessarily construed as a straightforward composition of individually meaningful parts, the meaning of the whole is associated with reliable arrangements of its constitutive elements. From this perspective the focus in morphology is shifted as in a Necker cube: instead of morphology being (solely) about the composition of complex word forms from smaller pieces, it is about complex surface word forms as representing types of configurations of elements and whole surface word forms as elements in a network of related word forms. As observed by Matthews (1991:204): 'words are not merely wholes made up of parts, but are themselves construable as parts with respect to systems of forms in which they participate'.

This emphasis on surface patterns of different sorts leads to a different set of research issues and questions. In particular, it becomes crucial to identify how complex words are organized into meaningful wholes without necessarily attributing meanings to identifiable parts, and how word forms are organized into structured networks (often referred to in terms of 'conjugations' and 'declensions'). In addition, it becomes natural to ask why the systems of organization cohere in the ways that they do, how such organization is learned, and whether the nature of the organization reflects learnability constraints, either specific to language or relevant in other learned domains as well. The paradigmatic/configurative perspective takes surface patterns seriously as entities that may facilitate learning, so the patterns represented by complex words and the patterns of organization among related words are, therefore, not the epiphenomenal result of representations and operations designed to produce individual words, as standardly assumed in syntagmatic/compositional approaches.

In such a WORD-AND-PARADIGM perspective, a surface word form is viewed as a RECOMBINANT GESTALT or configuration of recurrent elements (segmental or suprasegmental) that get distributed in complex words as members of a paradigm. As a consequence, the morphology of a language can be seen as an instance of a complex system, and the analysis of language more broadly begins to look like it can benefit from methods in other fields that study complex systems. Within language this parallels what has been described independently within the domain of speech sounds by Oudeyer (2006: 22) as the systematic reuse (and we would suggest SYSTEMIC reuse) of phonological distinctions: 'all languages have repertoires of gestures and combinations of gestures which are small in relation to the repertoires of syllables, and whose elements are systematically reused to make syllables'.

Likewise in morphology, as exemplified in detail below, the basic elements of words are used again and again in different configurations. Since configurations convey different meanings, this diminishes the need for bi-unique relations between forms and functions, rendering their syntagmatic arrangement and composition less essential to the morphological enterprise. The primary focus in the paradigmatic/configurational approach on surface words and their systematic alternants rather than on the identification of invariants responsible for deriving them recalls Anderson's (1985) insightful overview of Baudouin de Courtenay's and Kruszewski's theory of alternations in phonology and morphology. Describing the evolving conception of the phoneme in the works of these two linguists, Anderson argues that:

> It is worthwhile to notice, however, that the issue of such an invariant element arises most directly as a consequence of the need to deal with the systematic *variance* represented by the alternations. It is this systematic variation, with its fundamentally relational character, that language presents to us most directly. One way to organize this variation is to hypothesize underlying invariant units—indeed, judging from the history of the discipline, this is the most natural way for linguists to conceptualize such relations—but it should be borne in mind that this is not the only way to do so, or even the most transparent … for example, Saussure seems to have held a view of the phenomenon of variance and alternation that was much closer to an immediate account of the relations in question than to an account in terms of another kind of representation for linguistic forms, one given in terms of hypostatized invariants. (1985:68)

Crucially, the surface alternants were interpreted as participating in associative or paradigmatic networks of relations, requiring recognition of (networks of) whole words;

hence, this approach is exemplary of the paradigmatic/configurative perspective advanced here.

Crosslinguistically, inflectional morphology exhibits a spectacular range of variation with regard to the internal structure of individual words: languages differ with respect to the sizes as well as the particular inventories of morphosyntactic properties they distinguish and the formal strategies by which these properties are encoded. In this connection, Evans and Levinson (2009) argue that attention to similarities in terms of underlying uniformities across languages, such as those reflected in hypotheses about universal grammar, greatly understate observed crosslinguistic diversity; such claims also underestimate the value of variability as a guide for theory construction. They offer a provocative challenge to researchers in linguistics and the cognitive (neuro)sciences, namely, to recognize the consequences of how competing conceptualizations of linguistic phenomena encourage or discourage productive lines of inquiry. Most germane, they suggest that there are successful models in the biological sciences for dealing with variability that do not posit universals in order to address commonalities. Griffiths (2012:328) summarizes this view within evolutionary developmental biology:

> Comparative biology provides sophisticated ways to think about commonalities that underlie biological diversity. Bringing order to that diversity is not about identifying universal elements, but about finding order in the pattern of similarity and difference.

Evans and Levinson suggest that it is the range of surface diversity of languages themselves that provides the most instructive clues about the nature of human language. In effect, linguistic diversity is instructive because, among other reasons, it extends our notions of what is possible in human language. This way of seeing language has important parallels with the role of phenotypic variability in (ecological) developmental evolutionary biology, where variation is not seen as deviation from a canonical form, but as conditioned by multiple causally interleaving factors. Rarities are often more illuminating than what commonly occurs with respect to the possible shapes of forms.[2] Moreover, the suspicion ensues that the bounds on possible forms depend on constellations of interacting factors and cannot be construed as limitations on or as deviations from an idealized form.

When widespread crosslinguistic variability becomes a focal concern of linguistic inquiry, new questions arise that could not be previously formulated and new toolkits of analysis become relevant in order to provide answers. Within the domain of morphology, Evans and Levinson (2009:434) note that:

> much of the obvious typological difference between polysynthetic languages and more moderately synthetic languages like English or Russian needs to be taken at face value: the vast difference in morphological complexity is mirrored by differences in grammatical organization right through to the deepest levels of how meaning is organized.

While they propose several specific factors to account for how language-particular differences in morphological complexity may correspond to language-particular differences in grammatical organization, they neglect a crucial theoretical question. Are there general constraints on such exuberant diversity that obtain across different language systems? Recently, Carstairs-McCarthy (2010) has proposed a specific cognitive constraint on the shape and size of inflectional paradigms, namely VOCABULAR CLARITY. This constraint is argued to condition the possible shape of inflectional organization and is compatible with the existence of wildly complex paradigmatic organization, as long as it is met.

---

[2] See Alberch 1989 and Blumberg 2009 on rarities in biology, and Wohlgemuth & Cysouw 2010 for a discussion more directly relevant to language.

In this article, we argue for a more general constraining factor that is attested in other domains displaying patterned organization. Specifically, we argue that morphological systems reliably display low conditional entropy between the word forms in paradigms, in a sense to be made more precise in following sections. This constraint subsumes Carstairs-McCarthy's proposed constraint, rendering its effects as side-products of this more general organizing principle. Among other possible factors, morphological organization is guided and licensed by a simple constraint on viable, learnable pattern organization within both language and other domains of cognition.

Both across and within languages, the morphological complexity of words often serves the same communicative purposes as syntax. Partially agglutinative and partially fusional languages like Tundra Nenets or pervasively polysynthetic ones like Archi use morphological resources to encode meanings that are expressed by phrasal syntactic constructions in a language like English or Mandarin: for clausal predicators, these are properties such as pronominality, agreement, tense, aspect, and mood, among others. This is the empirical motivation for the guiding insight in lexical-functional grammar that morphology competes with syntax: grammars make use of two distinct ways to convey the same sorts of morphosyntactic properties (Bresnan 1998). This reflects the different strategies languages can employ to solve similar problems: there is no principled method nor any empirical motivation to convincingly reduce either strategy to the other.

The fact that inflectional morphology can encode such rich ensembles of information leads naturally to inquiries concerning the relative degrees of complexity within morphological systems: this is typically evaluated in terms of the number and the types of formal means by which such information is expressed. Most notably, some languages can have large inventories of morphosyntactic distinctions and highly articulated ways to formally express them, as in this West Greenlandic example (Fortescue 2002:264), where a concatenation of markers combines to produce a complex word.

(1)	aju-	nngit-	su-	liur-	vigi-	nnit-	tuar-	tu-	u-
	be.good-	NEG-	PART-	make-	have.as.place.of-	ANTIP-	all.the.time-	PART-	be-
	nngil-	aq							
	NEG-	3SG.IND							

'He is not (much of) a benefactor.'

Informally speaking, a language like West Greenlandic packages more information into words than English and does this by employing an impressive inventory of markers arranged in specific morphotactic orders relative to a lexical stem, as well as to one another.

Descriptive linguists often comprehensively catalogue the array of morphological markers and patterns in a given language or languages. Such data can be used for typological purposes to investigate the types of information that get encoded in words and the taxonomies of formal strategies for encoding this information, and for exploring issues such as why particular classes of markers tend to exhibit similar morphotactics. Theoretical linguists, by contrast, can use observed patterns of morphological packaging to infer the bounds on possible word structures in natural language. For example, one way to understand the evident variety in the syntagmatic and paradigmatic dimensions of inflectional morphology is to develop taxonomies of attested strategies for structuring words and delimiting the inventories of notions that they express. This can serve the goal of defining a crosslinguistic explanation for these particular (types of) morphosyntactic properties and their surface exponence. We refer to patterns found via this general cataloguing of properties and their surface exponence for words in all of their variety as the enumerative complexity or E-complexity of a morphological system.

As mentioned above, West Greenlandic words seem at an intuitive level more complex than the words of English: this can be reformulated as Greenlandic words displaying greater E-complexity than English words. In an obvious sense all theories must recognize the basic descriptive parameters of E-complexity, as these facts provide clues about the nature of morphology. A question arises, however, as to how illuminating such data can be with regard to the explanatory assumptions informing particular theoretical models: how does such data bear on the guiding assumptions in specific theories? In particular, are some types of patterns more informative than others, and what are the specific aspects of these patterns that provide the most insight into morphological systems? Moreover, how well do proposed explanatory principles and analytic tools in linguistic theories comport with those employed in other disciplines that have successfully explored complex phenomena? In other words, how well do linguistic theories fare with respect to cross-disciplinary parsimony in the conceptualization of their objects of analysis and the toolkits employed for analyzing them? In order to address these issues it is helpful to consider some specific properties of E-complexity.[3]

One salient dimension of E-complexity is the number and nature of inflection classes in a language. Anyone familiar with comparative morphological analysis knows that languages sometimes exhibit widespread allomorphy: this can produce almost kaleidoscopic combinatorics between roots or stems and morphosyntactic marking strategies. As a consequence, individual lexemes are often organized into large numbers of distinct inflection classes. The results of this can be extraordinary, as observed in Ackerman et al. 2009:55:

> … a typical Estonian noun paradigm contains some 30-odd forms, which exhibit patterns of variation that place the noun within anywhere between a half-dozen and a dozen major declension classes. It is implausible to assume that a speaker of Estonian will have encountered each form of every noun, so that native command of the language must involve the ability to generalize beyond direct experience.

Moreover, it was noted that the complexity of the Estonian system is dwarfed by languages such as Georgian and Archi, where estimates for the number of forms that verbs can have range from about 200 to 'more than one and a half million' (Kibrik 1998:467) respectively.

The complexity associated with allomorphy often only loosely correlates with systematic phonological or semantic conditions and frequently seems to serve no apparent communicative function. In this connection, all natural languages show a certain degree of what Baerman and colleagues (2010:2) call 'gratuitous' morphological complexity. Take, for example, the somewhat simplified Modern Greek nominal paradigms in Table 1.[4]

	SINGULAR				PLURAL			
CLASS	NOM	GEN	ACC	VOC	NOM	GEN	ACC	VOC
1	-os	-u	-on	-e	-i	-on	-us	-i
2	-s	-Ø	-Ø	-Ø	-es	-on	-es	-es
3	-Ø	-s	-Ø	-Ø	-es	-on	-es	-es
4	-Ø	-s	-Ø	-Ø	-is	-on	-is	-is
5	-o	-u	-o	-o	-a	-on	-a	-a
6	-Ø	-u	-Ø	-Ø	-a	-on	-a	-a
7	-os	-us	-os	-os	-i	-on	-i	-i
8	-Ø	-os	-Ø	-Ø	-a	-on	-a	-a

TABLE 1. Modern Greek nominal inflection classes (Ralli 1994, 2002).

Following Ralli (1994), Modern Greek has eight nominal declensions distinguished by the choice of endings for four cases and two numbers (setting aside stem alternations and stress patterns; see Sims 2010). Since the assignment of lexical items to particular

---

[3] For examples of efforts to identify and quantify E-complexity, see, for example, Juola 1998, 2007, Sampson et al. 2010, Moscoso del Prado Martín 2011.

[4] The Modern Greek paradigms as described by Ralli abstract away from many relevant complexities. For a more comprehensive description of the facts in the context of information-theoretic morphology, see Sims 2010, 2011, where the inclusion of greater detail also results in low conditional entropies.

declensions is essentially arbitrary, this adds a dimension of complexity to the inflectional system that serves no communicative purpose. It seems clear that this sort of morphological organization is much more complicated than is required by the mere number of morphosyntactic properties and forms needed to express them.

An E-complexity perspective leads to a number of fundamental questions facing morphological theory. Why would languages include morphological complexity whose only evident effect would seem to make a system harder to learn? Why would such systems, as largely the product of historical contingencies, often exhibit robustness that persists over centuries of linguistic evolution? Are there limits to such complexity, and if so, what constrains it?

While syntagmatic considerations (i.e. questions of word-internal structure and strategies for encoding variably sized inventories of morphosyntactic properties) have predominated in evaluating E-complexity, paradigmatic considerations of contrasts between words have tended to play a smaller role.[5] The latter perspective, however, highlights a crucial and recurring dimension in complex morphological systems, namely, the specification of what word form(s) could have appeared in a language, but did not: this identifies the nature of information associated with words in terms of their sets of alternative realizations. In other words, while, syntagmatically, each word itself conveys isolable information, there is another notion of information that no word can convey on its own: this is a notion of information that entails recognizing contrasts between (sets of) words. This notion of information derives from INFORMATION THEORY, and it permits us to ask a different type of question than is ordinarily addressed in syntagmatic approaches to morphology. It permits us to ask about the nature of the systems of organization that complex words find as their niche. Syntagmatics and E-complexity address the meaning and structure of complex words, while paradigmatics addresses the information associated with relations among words within morphological systems.

In this article, we suggest that insight into morphological organization is not obtained by addressing surface indicators of morphological complexity per se, but by examining, instead, the systemic organization underlying the surface patterns: what we refer to as the INTEGRATIVE COMPLEXITY, or I-COMPLEXITY, of a morphological system. As is shown below, constraints on systems of word shapes are best viewed in terms of I-complexity, providing a limit on E-complexity more broadly. This is in the service of the laudable goal of explaining observed restrictions on possible variety in natural language, which is identified as a major concern of linguistic theory in Cinque & Rizzi 2008.

This view of I-complexity provides a way to address the language diversity identified in Evans & Levinson 2009, while also suggesting that is not ‘without limit’ or restriction. According to the view developed here, I-complexity is a metric that reveals to what extent morphological systems are organized in ways that allow them to be learned and used by native speakers, irrespective of how complex words and paradigms may seem to be according to external measures.[6] Furthermore, we demonstrate that this notion of morphological organization is quantifiable in information-theoretic terms.[7]

---

[5] However, see Bybee 1985, Anderson 1992, Aronoff 1993, Stump 2001.

[6] I-complexity, in this sense, may provide a principle that facilitates the learning of complex morphology in accordance with Lupyan and Dale’s (2010) hypothesis about the relationship between population size and what we have referred to as E-complexity. They impute to early learners the capacity to learn more complex systems, but they do not provide a sense of how such complexity could be rendered learnable. Recent research (Stoll et al. 2012) demonstrates the utility of the types of entropy measures employed in Moscoso del Prado Martín et al. 2004 for explaining the order of acquisition of nouns versus verbs in Chintang, an Eastern Kiranti language spoken in Nepal.

[7] See, for example, Moscoso del Prado Martín 2003, Moscoso del Prado Martín et al. 2004, Hay & Baayen 2005, Blevins 2006, 2013b,c, Bonami & Boyé 2007, Milin et al. 2009, Bonami & Henri 2010.

One important requirement of any morphological system is that it must be possible for speakers to make accurate guesses about unknown forms of words based on exposure to known forms, despite the internal structural complexity of words and proliferations of conjugation/declension classes: speakers must generalize beyond their direct and limited experience of particular words to make likely inferences about unknown forms of that word within the system of paradigmatic relations characteristic of a specific language. This insight concerning the organization of paradigms in terms of reliable inferences or implications between words is expressed in traditional grammars via PRINCIPAL PARTS (see Stump & Finkel 2013 for a recent formalization) and more broadly with respect to the relevance of any word to other words in terms of implicative relations encoded in Wurzel's (1989) PARADIGM STRUCTURE CONSTRAINTS (Blevins 2013b, Bonami 2013).

We argue that morphological systems with low conditional entropy among related word forms in paradigms permit these crucial inferences to be made easily. The LOW ENTROPY CONJECTURE is the hypothesis that enumerative morphological complexity is effectively unrestricted, as long as the average conditional entropy, a measure of integrative complexity, is low (in a sense to be made more precise below).

On this view, it is misguided to search for grammatical principles that can predict restrictions on E-complexity. We demonstrate that when words and paradigms are considered as primary objects of analysis, morphological systems, which may be fiendishly complex from the perspective of a descriptive field linguist or lexicographer and utterly unpredictable from any imaginable principles of grammar, exhibit surprisingly low entropy, both absolutely and relative to what the entropy values could have been given the enumerative resources of particular grammars. Since entropy is effectively a measure of the reliability of guessing unknown forms on the basis of known ones, low entropy morphological systems are learnable and efficiently usable by native speakers: in a wholly straightforward sense this command of the morphological system represents native-speaker COMPETENCE, that is, speakers' knowledge of their language. Moreover, these patterns are rendered learnable by a general principle that is present in all patterned organization, irrespective of domain, namely conditional entropies that facilitate good guesses in the face of uncertainty: the uncertainty is tamed by the patterned organization provided by experience.[8]

As we progress it will become clear that the single measure of conditional entropy subsumes two independent strands of research in the literature on paradigm organization. The first concerns claims like the paradigm economy principle and its later variants, the no blur principle and the principle of synonymy avoidance (Carstairs-McCarthy 2010), and the second is the traditional reliance on principal parts, which comes from classical philology (Stump & Finkel 2007, 2013). While neither are generally assumed to relate to the other nor argued to derive from the same principle, we argue that both can be seen as descriptions of ways that profiles of E-complexity are demonstrably consistent with the maintenance of low entropy. Thus, they can be interpreted as special instances of this more general principle.

---

[8] We focus in this article on a speaker-centered perspective. Hearers, of course, face an analogous problem, namely, identifying the morphosyntactic feature set associated with a novel word form. The NAIVE DISCRIMINATIVE READER explored by, for example, Baayen and colleagues (2011) offers a probabilistic model of interpretation that is broadly compatible with what we describe here. It should also be noted that throughout we are assuming that speakers are familiar with the sets of morphosyntactic distinctions of their language. It is, of course, an entirely different question to identify how speakers arrive at this knowledge. We are presently investigating this problem from the perspective of the relation between information theory and discriminative learning (e.g. Ramscar & Dye 2009). Bonami (2013) explores this issue using a variant of Albright and Hayes's (2003) minimal generalization algorithm.

We begin by introducing information theory and the way that conditional entropy can be calculated in relatively simple paradigms. We then turn to a comparative analysis of languages displaying various degrees of enumerative complexity in §3, demonstrating that, despite their large surface differences, they all share the property of low average conditional entropy. Section 4 provides a detailed look at a test case, a language that displays stunning E-complexity but, surprisingly, a level of I-complexity in line with that, of apparently much simpler languages. We also provide independent evidence for the entropy-guided organization of the paradigms in this language and develop a Monte Carlo simulation that demonstrates that the paradigm organization we actually find corresponds to an adaptive solution that yields low entropy for all of the morphosyntactic properties and formal encodings in this language. Concluding remarks summarize our results and elaborate on their consequences for the analysis of morphological systems.

2. CONDITIONAL ENTROPY. In classical paradigm-based models of morphology, a morphological system is represented via two distinct components: a set of exemplary full paradigms that exhibit the inflectional classes of a language, and sets of diagnostic principal parts that can be used to deduce which inflectional class a given lexeme belongs to.[9] By hypothesis, speakers may memorize complete paradigms for frequent lexemes, but for infrequent lexemes speakers must produce word forms by analogic extension of known exemplars. Given the right word forms of a novel lexeme, word-and-paradigm models provide a general strategy for filling in the rest of the paradigm. This exploits the implicational structure of inflectional systems in a way that can be quite effective.[10] In general, a small set of diagnostic principal parts is often sufficient to identify the inflectional class of a lexeme and thus to accurately predict other word forms of the lexeme. Paradigm-based models also reflect a measure of E-complexity: languages with a greater number of possible exponents, inflectional classes, and principal parts will require more word forms to be memorized by the language user (and recorded by the lexicographer).

From the point of view of the fluent language user, however, this is an artificial measure of complexity. While speakers of morphologically complex languages do often have to produce word forms that they have never heard before, they rarely have to predict all forms of a given lexeme. On the contrary, speakers must produce some subset of the complete paradigm of a lexeme given knowledge of some other subset, a task that often will not require completely resolving a lexeme's inflectional class membership. However, speakers have no guarantee that they will have been exposed to the most relevant or diagnostic principal parts of a novel lexeme. Patterns of implicational relations among word forms within paradigms can be interpreted as providing speakers with a means for carrying out these predictions with incomplete information.

In order to assess the implicational relations among word forms, we use the information-theoretic notion of ENTROPY as the measure of uncertainty or predictability. This permits us to quantify 'prediction' as a change in uncertainty, or information entropy (Shannon 1948). Suppose we are given a random variable X that can take on one of a set of alternative values x_1, x_2, ... , x_n with corresponding probability $p(x_1)$, $p(x_2)$, ... ,

---

[9] See, for example, Matthews 1974, Stump 2001, Ackerman & Stump 2004, Trosterud 2004, Blevins 2005, 2007, Albright 2009.

[10] See recent work on the role of analogy-based motivations for paradigm change in Albright & Hayes 2003, Albright 2005, 2009. Though the present research is limited to the characterization of synchronic systems, the exploration of how such systems arrive at their present states is, of course, a crucial diachronic and psycholinguistic concern. In this connection, work of Albright and Hayes (2003) on structured and variegated similarity, as well as their uses in the probability-based model of Albright 2009, is particularly useful.

$p(x_n)$. Then, the amount of uncertainty in X, or, alternatively, the degree of surprise on learning the value of X, is the entropy $H(X)$.

$$(2)\quad H(X) = -\sum_i p(x_i) \log_2 p(x_i)$$

The entropy $H(X)$ is the weighted average of the SURPRISAL $-\log_2 p(x_i)$ for each possible outcome x_i. The surprisal is a measure of the amount of information expressed by a particular outcome measured in bits, where one bit is the information in a choice between two equally probable outcomes. Outcomes that are less probable (and therefore harder to predict) have higher surprisal. Specifically, surprisal is zero bits for outcomes that always occur ($p(x) = 1$) and approaches ∞ for very unlikely events (as $p(x)$ approaches zero). The more choices there are in a given domain and the more evenly distributed the probability of each particular occurrence, the greater the uncertainty or surprise there is (on average) that a particular choice among competitors will be made, and, hence, the greater the entropy. Conversely, choices with only a few possible outcomes or with one or two highly probable outcomes and lots of rare exceptions have a low entropy. Furthermore, INFORMATION in this context is anything that resolves uncertainty. If there is a high degree of uncertainty in the value of a random variable, then we will (on average) be highly surprised when we find out what the value actually is, and a message conveying that value will contain a lot of information.

For example, the entropy of a coin flip as resulting in either heads or tails is one bit; there is equal probability for an outcome of either heads and tails.

$$\begin{aligned}(3)\quad H(X) &= -\sum_i p(x_i) \log_2 p(x_i)\\ &= -(p(\text{heads}) \times \log_2 p(\text{heads}) + p(\text{tails}) \times \log_2 p(\text{tails}))\\ &= -(0.5 \times \log_2 0.5 + 0.5 \times \log_2 0.5)\\ &= 1 \text{ bit}\end{aligned}$$

So, it will require (on average) one bit to encode a message reporting the result of a coin flip. The entropy of a coin rigged to always come up heads, by contrast, is zero bits—there is no uncertainty in the outcome (assume $0 \log_2 0 = 0$).

$$\begin{aligned}(4)\quad H(X) &= -\sum_i p(x_i) \log_2 p(x_i)\\ &= -(p(\text{heads}) \times \log_2 p(\text{heads}) + p(\text{tails}) \times \log_2 p(\text{tails}))\\ &= -(1.0 \times \log_2 1.0 + 0.0 \times \log_2 0.0)\\ &= 0 \text{ bits}\end{aligned}$$

A message reporting the result of this rigged coin will require zero bits: since the coin always comes up heads, we know the outcome before the coin is even flipped and no message will provide any additional information.

With this as background we can now return to the partial Modern Greek nominal paradigms in Table 1 to quantify the uncertainty among the nominal types. Suppose we want to represent the inflection class membership of an arbitrary lexeme. This is, for instance, the problem faced by a lexicographer preparing a dictionary of the language. If D is the set of declensions for a particular paradigm, the probability (assuming all declensions are equally likely) of an arbitrary lexeme belonging to a particular paradigm d is as follows.

$$(5)\quad P(d) = \frac{1}{|D|}$$

Since in the Modern Greek example there are eight distinct classes, the probability of any lexeme belonging to any one class would be 1/8. We could represent a lexeme's declension as a choice among eight equally likely alternatives, which thus has an entropy of $-\log_2 8 = 3$ bits. This is the DECLENSION ENTROPY $H(D)$, the average information required to record the inflection class membership of a lexeme.

Of course, not all inflection classes are equally likely. In general, some classes will have more members than others, and a randomly selected lexeme is more likely to be a member of a class with many members. Let $F_{\text{typ}}(d)$ be the TYPE FREQUENCY of declension d, that is, the number of lexemes that are members of that class. Then, in general, the probability of a declension d is as given in 6.

$$(6)\quad P(d) = \frac{F_{\text{typ}}(d)}{\sum_{d \in D} F_{\text{typ}}(d)}$$

That is, the probability of a randomly selected word being in declension d is just the number of lexemes that actually are in declension d divided by the sum of the lexeme counts for all declensions (which in turn is just the total number of lexemes in the relevant vocabulary). Factoring type frequency into our calculation of declension entropy can only reduce our estimate, sometimes substantially. If, for example, half of all lexemes fall into one class, with the remaining half evenly distributed among the remaining eight classes, then our estimate of the declension entropy is reduced from three bits to 2.4 bits.

In many cases, inflectional class membership is also at least partly predictable from external factors, such as the phonological shape or lexical gender of the root. Any information that helps speakers predict the realization of a word form can only reduce the entropy. Since for present purposes we ignore these factors, the entropy values we present are upper bounds, recognizing that if all factors are taken into account the actual entropies will likely be much lower.

Recording the declension of an arbitrary noun lexeme (the problem faced by our hypothetical lexicographer) is more difficult than the problem faced by a speaker; a dictionary might be maximally useful if it provides complete entries for all of the inflected forms of a (class of) lexeme (e.g. Holman 1984), while speakers need only produce one single form (in any particular context). When we look at individual paradigm cells we find much less uncertainty than the declension entropy would lead us to expect. While there are eight declensions in Table 1, there are at most five distinct realizations for any individual cell, and in each instance some realizations are more likely than others.

Let $D_{c=r}$ be the set of declensions for which the paradigm cell c has the formal realization r. Then the probability $P_c(r)$ of a paradigm cell c of a particular lexeme having the realization r is the probability of that lexeme belonging to one of the declensions in $D_{c=r}$.

$$(7)\quad P_c(r) = \sum_{d \in D_{c=r}} P(d)$$

The entropy of this distribution is the PARADIGM CELL ENTROPY $H(c)$, the uncertainty in the realization for a paradigm cell c. Carrying out the necessary calculations for the Modern Greek sample paradigms, we get the values in 8.

(8)

c	NOM.SG	GEN.SG	ACC.SG	VOC.SG	NOM.PL	GEN.PL	ACC.PL	VOC.PL	AVG
$H(c)$	1.750	2.156	1.549	1.549	1.906	0.000	2.156	1.906	1.621

Note that the paradigm cell entropy varies across the paradigm cells. The GEN.PL has only one possible realization and an entropy of zero bits, but even the most diverse cells have a lower entropy (at 2.156 bits, assuming uniform declension probabilities) than the declension system as a whole. The average entropy across all cells is 1.621 bits; this average is a measure of how difficult it is for a speaker to guess the realization of any one word form of any particular lexeme in the absence of any information about that lexeme's declension. An entropy of 1.621 bits is equivalent to selecting among only $2^{1.621} \approx 3$ equally likely alternatives. That is, the Modern Greek noun paradigms in Table 1 fall into eight declensions, but selecting the realization for a particular word form of a lexeme is as difficult as a choice among about three equally likely alternatives.

Guessing the realization of a single word form is quite a bit easier than guessing the declension of a lexeme. But even this overstates the complexity of the system, as speakers must have some information about the lexeme in order to know that the lexeme even exists. At a minimum, speakers will know at least one word form of a lexeme for which they wish to produce a novel word form. A careful look reveals that the structure of the Modern Greek nominal paradigm is such that there is a high degree of interpredictability among word forms. Take the two most complex word forms, the GEN.SG and the ACC.PL. Since each has five possible forms, there is a potential for 5 × 5 pairings. But only seven are actually attested in the lexicon, as seen in Figure 1.

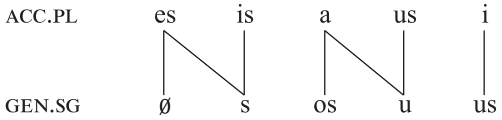

FIGURE 1. Attested pairings of ACC.PL and GEN.SG realizations in Modern Greek.

The GEN.SG is not fully interpredictable with the ACC.PL in the sense that one form is reliably deducible from the other by a rule. However, knowing one word form narrows down the possibilities for the other, or, to put it another way, one word form provides useful information about the other.

To quantify the predictability of one form given the other, we can measure the size of the surprise associated with these forms using CONDITIONAL ENTROPY $H(Y|X)$, the uncertainty in the value of Y given that we already know the value of X.

$$(9)\quad \begin{aligned} H(Y|X) &= H(X,Y) - H(X) \\ &= \sum_{x\in X}\sum_{y\in Y} P(x,y)\log_2 P(y|x) \end{aligned}$$

The smaller $H(Y|X)$ is, the more predictable Y is on the basis of X: that is, the less surprised one is that Y is selected given knowledge of X. In the case where X completely determines Y, the conditional entropy $H(Y|X)$ is zero bits: given the value of X, there is no question remaining as to what the value of Y is. By contrast, if X gives us no information about Y at all, the conditional entropy $H(Y|X)$ is equal to $H(Y)$: given the value of X, we are just as uncertain about the value of Y as we would be without knowing X at all.

Above we defined $P_c(r)$, the probability that paradigm cell c of a lexeme has the realization r. We can easily generalize that to the joint probability of two cells c_1 and c_2 having the realizations r_1 and r_2 respectively.

$$(10)\quad P_{c_1,c_2}(r_1,r_2) = \sum_{d\in D_{c_1=r_1\wedge c_2=r_2}} P(d)$$

To quantify paradigm cell interpredictability in terms of conditional entropy, we can define the conditional probability of a realization given another realization of a cell in the same lexeme's paradigm.

$$(11)\quad P_{c_1}(r_1|c_2=r_2) = \frac{P_{c_1,c_2}(r_1,r_2)}{P_{c_2}(r_2)}$$

With this background, the conditional entropy $H(c_1|c_2)$ of a cell c_1 given knowledge of the realization of c_2 for a particular lexeme is as follows.

$$(12)\quad H(c_1|c_2) = \sum_{r_1}\sum_{r_2} P_{c_1}(r_1)\, P_{c_2}(r_2) \log_2 P_{c_1}(r_1|c_2 = r_2)$$

In the case of the Modern Greek forms in Fig. 1, if we know that the ACC.PL of a lexeme is in *-i*, then there is no uncertainty as to the GEN.SG: it must be in *-us*. That is, $H(\text{GEN.SG}|\text{ACC.PL} = \text{-}i)$ is zero. If, though, the ACC.PL is in *-a*, there are two possibilities for the GEN.SG. In two of the three declensions where the ACC.PL is in *-a*, the GEN.SG is in *-o*, and in one the GEN.SG is in Ø. Therefore the conditional entropy is as follows.

$$(13)\quad H(\text{GEN.SG}|\text{ACC.PL} = \text{-}a) = -\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right)$$
$$= 0.918 \text{ bits}$$

Averaging across each of the possible realizations for the ACC.PL, we get the conditional entropy in 14.

$$(14)\quad H(\text{GEN.SG}|\text{ACC.PL}) = 0.594 \text{ bits}$$

In other words, while guessing the GEN.SG of a lexeme is as difficult as a choice among $2^{2.156} = 4.312$ alternatives, guessing the GEN.SG on the basis of the ACC.PL requires (on average) a choice among only $2^{0.594} = 1.188$ alternatives.

The conditional entropy is a measure of the difficulty of solving one instance of what Ackerman et al. 2009 refers to as the paradigm cell filling problem: predicting a specific unknown word form from a specific known word form. A complete table of pairwise conditional entropies for our fragment of Modern Greek is given in Table 2.

$H(col\|row)$	NOM.SG	GEN.SG	ACC.SG	VOC.SG	NOM.PL	GEN.PL	ACC.PL	VOC.PL	E[row]
NOM.SG	—	1.000	0.250	0.250	0.750	0.000	1.000	0.750	0.571
GEN.SG	0.594	—	0.594	0.594	0.594	0.000	0.594	0.594	0.509
ACC.SG	0.451	1.201	—	0.000	0.951	0.000	0.951	0.951	0.644
VOC.SG	0.451	1.201	0.000	—	0.951	0.000	0.951	0.951	0.644
NOM.PL	0.594	0.844	0.594	0.591	—	0.000	0.250	0.000	0.411
GEN.PL	1.750	2.156	1.549	1.549	1.906	—	2.156	1.906	1.853
ACC.PL	0.594	0.594	0.344	0.344	0.000	0.000	—	0.000	0.268
VOC.PL	0.594	0.844	0.594	0.594	0.000	0.000	0.250	—	0.411
E[col]	0.719	1.120	0.561	0.561	0.736	0.000	0.879	0.736	0.664

Table 2. Conditional entropies for Modern Greek paradigms in Table 1.

These values tell us how difficult it is to guess one particular word form on the basis of one other particular word form. In general, however, we cannot predict which forms a speaker will generalize from or to. This will depend on the cell probability $P(c)$, the probability that a randomly selected word form is some lexeme's realization of cell c. In the simplest case we can assume that all cells are equally likely, so if C is the set of cells in a paradigm, then the probability is as follows.

$$(15)\quad P(c) = \frac{1}{|C|}$$

Or, we could estimate $P(c)$ from the TOKEN FREQUENCY $F_{\text{tok}}(c)$ of the cell c in a representative corpus.

$$(16)\quad P(c) = \frac{F_{\text{tok}}(c)}{\sum_{c \in C} F_{\text{tok}}(c)}$$

Given $P(c)$, the expected values $E[\text{col} = c_1]$ and $E[\text{row} = c_2]$ are the average uncertainty in guessing the form of some cell c_1 or guessing based on the form of cell c_2 (respectively).

(17) $$E[\mathrm{col} = c_1] = \sum_{c_2} P(c_2)\, H(c_1|c_2)$$
$$E[\mathrm{row} = c_2] = \sum_{c_1} P(c_1)\, H(c_1|c_2)$$

For example, on average the uncertainty in the NOM.SG form of a word given the realization of another randomly selected cell is 0.719 bits, and guessing a randomly selected other word form on the basis of the NOM.SG is 0.571 bits. The average across all possible pairs of word forms is the overall average conditional entropy.

Averaging across all possible pairs of word forms, the overall AVERAGE CONDITIONAL ENTROPY $H(P)$ is as follows.[11]

(18) $$\begin{aligned} H(P) &= \sum_{c_1} P(c_1)\, E[\mathrm{col} = c_1] \\ &= \sum_{c_2} P(c_2)\, E[\mathrm{row} = c_2] \\ &= \sum_{c_1} \sum_{c_2} H(c_1|c_2) \end{aligned}$$

For our simplified Greek example, $H(P)$ is 0.644 bits, equivalent to a choice among only $2^{0.644} = 1.56$ equally likely declensions. That is, while this fragment of Modern Greek has eight declensions from the point of view of a lexicographer trying to describe the language, for a speaker trying to use the system it has on average only slightly more than one: this is the I-complexity of this paradigm in Modern Greek. Accordingly, inflectional systems in the form of Table 1 have the potential to greatly overstate the apparent complexity of a morphological system.

3. A SMALL SAMPLE OF LANGUAGES.

As previously mentioned, words and paradigms often vary substantially in their apparent E-complexity. To illustrate this variability we have selected a small sample of languages that have formed the basis for previous analyses by others in different frameworks.[12] It is important to note that these languages were not chosen because they were antecedently identified as confirming the low entropy conjecture, but rather because they represented ten geographically and, for the most part, genetically diverse languages whose previous analyses had not demonstrated the sort of fundamental relatedness between them that is possible in terms of I-complexity. As a consequence, the results from this sample regarding conditional entropy suggest the value of expanding genuine typological inquiry to a broader set of languages with more complete

---

[11] There is an important dimension of conditional entropy calculation that we do not address in this article. While average conditional entropy provides an average arising across all forms, it is clear from Table 2 that some conditional entropies between pairs of cells are considerably lower than those between other pairs. This raises an empirical question as to whether there may be general principles affecting the organization of paradigms such that certain cells tend to be more predictive and others more predictable. A question arises concerning whether frequencies play a role in such organization, such that more frequent forms tend to be more predictive than less frequent, hence predictable, forms. For example, since certain inflected forms are more frequent, that is, more likely to be encountered, than other forms, does the patterned structure reflective of average conditional entropy reveal any preferences for the former to be more predictive and the latter to be more predictable? As previously noted, the reader is directed to Sims 2010, 2011 for a careful and detailed examination of Modern Greek that addresses these issues as well as the role of type and token frequencies in entropy calculations; the latter factors are also explored in Bonami & Henri 2010.

[12] The calculations below are based on partial paradigms from Ralli 1994 (Greek), Stump & Finkel 2007 (Fur, Kwerba, Ngiti), Hein & Müller 2009 (Amele, Arapesh), Baerman et al. 2010 (Burmeso, Nuer, Russian), and Baerman & Corbett 2010 (Mazatec). All of the data in these sources were taken from primary grammar sources, which we have also consulted and, in some instances, have used to guide our analysis in different directions than were taken in previous analyses. For details, see the appendix.

paradigms. Given the acknowledged limitations of this analysis, Table 3 contains the relevant information for partial paradigms from these ten languages.

LANGUAGE	CELLS	REALIZATIONS	MAX REALIZATIONS	DECL CLASSES	DECL ENTROPY	PARADIGM ENTROPY	AVG ENTROPY
Amele	3	30	14	24	4.585	1.105	2.882
Arapesh	2	41	26	26	4.700	0.630	4.071
Burmeso	12	6	2	2	1.000	0.000	1.000
Fur	12	50	10	19	4.248	0.517	2.395
Greek	8	12	5	8	3.000	0.644	1.621
Kwerba	12	9	4	4	2.000	0.428	0.864
Mazatec	6	356	94	109	6.768	0.709	4.920
Ngiti	16	7	5	10	3.322	0.484	1.937
Nuer	6	3	2	16	4.000	0.750	0.778
Russian	12	14	3	4	2.000	0.538	0.911

TABLE 3. Paradigm entropies.

The columns in this table reflect the basic distinctions with regard to the essential elements of paradigms and the different calculations of entropy presented above for Greek. The column labeled CELLS refers to the multidimensional morphosyntactic property sets that are occupied by the forms referred to in the column labeled REALIZATIONS: the former corresponds to what Ackerman and Stump (2004) and Stump (2006) refer to as content paradigms and the latter to form paradigms. For example, Amele has three distinct cells associated with morphosyntactic properties for possessed nominals (1SG, 2SG, and 3SG), but this small number of distinctions is formally expressed in thirty different ways. The column labeled MAX REALIZATIONS refers to the maximum number of distinct forms associated with a single cell. That is, for Amele, while there are thirty separate realizations, there is some single cell (namely, 3SG) that is realized by fourteen of these realizations. This would be the predicted limit of inflectional classes by the PARADIGM ECONOMY PRINCIPLE (see further discussion below). The fifth column lists the number of distinct declension classes that any given lexeme may belong to. For Amele, we see that there are twenty-four separate classes that encode the three cells and their thirty different possible realizations.

Column 6 identifies the DECLENSION ENTROPY for the relevant paradigms in each language. Recall that this measures the difficulty of guessing declension class assignment for a randomly selected lexeme, independent of knowledge about any of its specific forms. So, for Amele, this yields a declension entropy of 4.585. This reflects the informal intuition that a good guess in this domain is hard, since any random lexeme could belong to one of twenty-four separate classes. Column 7 provides the PARADIGM CELL ENTROPY, the uncertainty associated with the simpler task of guessing the realization of a lexeme as expressed by a single cell. Since any given cell is associated with a relatively small number of forms, this should be easier (and can never be harder) than correctly guessing the declension class. Accordingly, the entropies are lower in this column than in the sixth. In Amele, for example, there are fewer maximum realizations associated with any single cell (fourteen) than there are declension classes (twenty-four).

Finally, the last column provides the AVERAGE CONDITIONAL ENTROPY, the quantity of most relevance to the present study. In effect, the preceding two columns provided a measure for calculations of behaviors that are unrealistic for language speakers: native speakers are not confronted with situations where it is necessary to guess declension class membership or the cell for particular word forms of novel lexemes without having previously encountered some actual word form of that lexeme. This means that the speaker's task is more concrete and determinate than these more abstract considerations. Given that the speaker knows some specific form of a word, one can calculate the conditional entropy of how difficult it is to guess another specific form of that word. Thus, as described in §2, the eighth column provides the average conditional entropy for the relevant inflectional system of a language: this represents the average of all of the conditional entropies that arise when one form of a word is used to predict another form of that word. The entropies in this column are the lowest of those identified in Table 3, suggesting the value of particular word forms for reliably predicting related word forms. The average conditional entropy can never be larger than the declension entropy: guessing a lexeme's declension on the basis of a single word form can never be harder than guessing the declension with no information at all. What is notable is that for all of the languages in Table 3 the average conditional entropy is dramatically lower than either the average entropy or the declension entropy, and, more importantly, is uniformly low across a range of languages with dramatically differing degrees of morphological E-complexity.

While the pattern that emerges is very suggestive, these results must, of course, be considered with several caveats in mind. Obviously, a sample of ten languages reflects only a small range of the variation found in natural languages. Moreover, within each language we have only analyzed partial paradigms. Though a more representative and comprehensive sample is desirable, however, we have selected languages and paradigms with varying degrees of E-complexity as test cases for the low entropy conjecture. Finally, our entropy calculations are based on the descriptions of morphological systems culled from written grammars. These descriptions often omit crucial details, and, more importantly for determining the veridical I-complexities of inflectional systems, they do not convey the relative type or token frequencies associated with inflectional classes. Research that begins to address these issues through more detailed and direct statistical analyses of corpora (e.g. Bonami & Henri 2010, Malouf & Ackerman 2011, Sims 2011, Bonami 2013) appears to further validate the utility of average conditional entropy.[13] But given the novelty of the present program, additional work is required before the low entropy conjecture can graduate to a hypothesis or even a principle of paradigm organization.

Returning to Table 3, the only paradigm with an entropy greater than one bit is the Amele possessive inflection, and this, it appears, is based on a descriptive overstatement of the paradigm's complexity. Amele is described by Roberts (1987) as having thirty-one different classes of possessive suffixes plus a postpositional marking strategy. Using this description, Hein and Müller (2009) argue that factoring out phonologically predictable alternations reduces this to twenty-three suffixed classes plus the postpositional strategy. But, by Roberts's (1987) description, the suffixed classes (many of which have a single member) apply only to a closed class of 109 inalienably possessed nouns, and the choice of suffixes is largely determined by a set of not-quite-categorical semantic and phonolog-

---

[13] Baerman (2012) presents an analysis of a larger set of Nuer suffix classes which, by his calculations, have an average entropy of 1.24 bits and an average conditional entropy of 1.04 bits, which is quite a bit higher than the other languages in our small sample. However, he argues that the Nuer suffix system can only be properly understood in terms of its interaction with a system of stem alternations. We would conjecture that, as in Mazatec, the average conditional entropy of the joint stem alternation-suffix system would be lower than that of the suffixes taken alone. Baerman also reports results for Latvian nominal paradigms with an average entropy of 1.56 bits and an average conditional entropy of 0.66 bits, which is in the same range as the paradigms we have discussed here.

ical generalizations. Naturally, any statistical, phonological, or semantic patterns such as these that can be used to partially predict word forms will lower the entropy. So, the figures cited in Table 3 are upper bounds, and the actual entropies would be much closer to zero if all factors could be accounted for.[14] Crucially, however, marking of inalienable possession is a domain where the forms are likely to be memorized, and hence, entropy-oriented organization that facilitates good guessing is irrelevant. In fact, it appears that the postposition strategy is the only truly productive marking, and it should be quite simple to reliably guess novel forms employing this strategy, modulo knowledge of other relevant factors. The average conditional entropy values in Table 3 provide an empirical basis for the low entropy conjecture: inflectional paradigms appear to be organized in such a way that their average conditional entropies are relatively low. While one might suspect that this pattern is an artifactual result derived from the use of a small sample with incomplete paradigms, irrespective of obvious E-complexity differences among the languages, further reflection suggests that such results conform to intuitively sensible organization; in fact, it seems that the greater the E-complexity of a system, the more important organizing principles are. After all, languages must somehow be simple enough to permit productive use by reliable extrapolation to unknown forms.

Empirical evidence suggests that there are a number of alternative strategies by which low entropy can be achieved and maintained in the languages of the world. Perhaps the most straightforward way that low entropies are produced is for a language to possess a small repertoire of morphosyntactic properties, a small inventory of exponents, and few inflectional/declension classes. For example, consider Burmeso and Russian in Table 3. Burmeso has twelve cells, six exponents, and only two declension classes. Russian has twelve cells, fourteen exponents, and four declension classes. The Burmeso pattern yields an average conditional entropy of zero bits, while in Russian this figure is 0.538 bits. These systems are evidently simple enough from an E-complexity perspective that there is no need to organize their systems in terms of implicational relations.[15] Further evidence for the diminished importance of implicational relations in Russian will come from the computational experiment described in §4.

A second strategy is associated with two theoretical claims proposed by Carstairs-McCarthy. These are the PARADIGM ECONOMY (PE) PRINCIPLE (Carstairs 1987) and the NO BLUR PRINCIPLE (Carstairs-McCarthy 1994, 2010).

> … the PE Principle states that there can be no more inflectional paradigms for any word class in any language than there are distinct 'rival' inflectional realizations available for the morphosyntactic property-combination where the largest number of rivals compete. (Carstairs-McCarthy 1991:222)

In other words, there should be no more classes than the largest number of affixal allomorphs associated with a single cell.

For example, in the fragment of Modern Greek presented in Table 1, the genitive singular has six rival allomorphs competing to express this feature set. Since this is the largest inventory of rival forms, PE predicts that there can be no more than six inflec-

---

[14] As a referee points out, additional information is only guaranteed to lower the entropy given a specific inventory of classes. For a given inflectional class system, adding information can only reduce the entropy, but changing the number or structure of the inflection classes may increase the entropy. Not surprisingly, the analysis one assumes makes a big difference. This is an issue explored in detail in, for example, Bonami 2013 and Stump & Finkel 2013.

[15] As observed by a referee, the lack of necessity to construe a system such as Russian in terms of implicational relations does not mean that it is not pedagogically useful to do so, as evidenced by teaching and reference grammars of this language.

tional classes. It is evident, however, that Modern Greek as presented has eight declension classes. It is easily imaginable that individuating and enumerating inflection classes in any given language might be problematic, with the consequence that determining the number of classes is less obvious than it may seem. Carstairs-McCarthy, in fact, proposes various codicils and exemptions to PE that complicate this task, while trying to address exceptions: this obtains for the no blur principle as well.[16] However these issues are resolved, while Modern Greek may or may not have more than the predicted number of inflection classes, there is no question that it has fewer inflection classes than it could. If all exponents were combined in every combination to form an inflection class, Modern Greek could have as many as $4 \times 6 \times 4 \times 4 \times 6 \times 1 \times 5 \times 4$ or 46,080 distinct classes. So, PE is certainly more right than it is wrong.

In general, Carstairs-McCarthy's earlier proposals are based on the counting and combinatorics of surface forms in order to account for the E-complexity of a system: what is the relation of the inventory of allomorphs to the number of inflection/declension classes? In effect, Carstairs-McCarthy's contribution in that work was to provide a descriptive principle that constrains the number of potential classes to a size that better resembles the profile attested in natural language systems than might be found if there were no principle at all.[17] On the one hand, it must be observed that the effectiveness of these principles, to whatever degree they work, itself stands in need of explanation. In other words, it is not clear why such principles should exist or operate as they do, nor, as observed by a referee, precisely how they are intended to operate. Carstairs-McCarthy (2010) explores the source of the no blur principle's effects for somewhat straightforward departures from expectation in terms of cognitive principles such as SYNONYMY AVOIDANCE.[18] Independent of the role such cognitive principles may play in a given language, however, it is evident that the systems he examines all display low entropy: low entropy provides a unifying explanation that is affected by the specific factors such as those arising from cognitive principles. As a consequence, both simple systems such as Burmeso and Russian and the more complex systems he analyzes are explicable in terms of a single measure, namely, low entropy, irrespective of the particular factors that produce this result.

The same can be said for a third strategy for lowering entropy, specifically an appeal to principal parts, as in Stump & Finkel 2009. The basic motivation guiding the search for principal parts is to identify a small set of forms that are diagnostic of (inflection or declension) class membership for lexemes. On traditional assumptions principal parts reliably predict class membership, while other forms do not have diagnostic value: knowledge of principal parts facilitates correct inferences to novel forms of known lexemes. Recall that in the Modern Greek paradigms in Fig. 1, knowing that the ACC.PL of a lexeme is in *-i* predicts that the GEN.SG must be in *-us*.

While this is a classic case of the role of the implicational organization of morphological systems,[19] the conventional focus on fully diagnostic forms misses the essential insight that guides the present proposal: all word forms are diagnostic to some degree and

---

[16] The exceptions largely involve not counting inflection classes that, roughly speaking, do not increase the average conditional entropy; see the debate regarding the value of the no blur principle in Stump 2005 and Carstairs-McCarthy 2010.

[17] In more recent work, Carstairs-McCarthy (2010) has turned to focus on what we describe here as I-complexity. Also, see Müller 2007 for effort in much the same spirit as the paradigm economy principle to restrict logical limits to empirically attested ones.

[18] See Clark 1993 on the principle of contrast in first language acquisition.

[19] See Wurzel 1989.

contribute in some measure to implicational networks of relatedness between words. On our account, crucially, all forms are potentially associated with some degree of diagnostic value, since all words provide information to some degree. Returning again to the Modern Greek example, knowing that the ACC.PL of a lexeme is in *-a* does not predict the GEN.SG, but it does narrow down the possibilities.[20]

In this sense, traditional principal parts are simply those forms that analysts have reified into fundamental organizing entities. While some forms in some languages may be privileged with respect to their predictive role in such networks, this is not a necessary condition for I-complexity. In this sense, the construct 'principal part' is construable as the most obvious instance of the diagnostic value of forms; that is, these are forms that participate in the lowest conditional entropy relations. Thus, the organization of paradigms around a small ensemble of principal parts is merely one way that I-complexity can be achieved.

All three of the strategies that we have reviewed share a single property: they all achieve low entropy. This is, of course, striking, since it is not predicted a priori that morphological organization in languages with relatively impoverished E-complexity, languages that motivated the PE principle and the no blur principle, and languages that motivated an appeal to principal parts could be subsumed under a single descriptive generalization, specifically, the low entropy conjecture. This is consistent with the hypothesis that all inflectional paradigm organization will be similarly constrained. That is, we expect even large departures from the predictions of the no blur principle, and the bounds of identifiable principal parts should be consistent with the low entropy conjecture. In §4 we provide a small case study in how this measure accounts for a language with extraordinary E-complexity, which has been claimed to challenge the comprehensive coverage attributed to low entropy (Baerman & Corbett 2010).

4. A SURPRISINGLY SIMPLE EXAMPLE. In Chiquihuitlán Mazatec (CqM), verbs are marked for person and aspect by a combination of tones, final vowel, and stem formative (Jamieson 1982, Capen 1996, Baerman & Corbett 2010). This leads to an inflectional system that appears to be characterized by an extremely high degree of irregularity and suppletion, with few apparent implicational relations. Consider for example the (partial) paradigm in 19.

(19) Positive paradigm for *ba*^{3}*se*^{2} 'remember' (Jamieson 1982:166)

	NEUTRAL		INCOMPLETIVE	
	SG	PL	SG	PL
1INCL		ča^{2}sẽ^2		ča^{2}sẽ^{42}
1	ba^{3}sæ^1	ča^{2}sĩ^{24}	kua^{3}sæ^1	ča^{4}sĩ^{24}
2	ča^{2}se^2	ča^{2}sũ^2	ča^{4}se^2	ča^{4}sũ^2
3	ba^{3}se^2		kua^{4}se^2	

There is no obvious stem, as the only piece that recurs in all of these word forms is the consonant *-s-*. Instead, the verb glossed 'remember' is distinguished from other verbs by its membership in an inflectional class in each of three distinct cross-cutting dimensions: tone pattern, final vowel, and stem formative. In particular, *ba*^{3}*se*^{2} 'remember' is a member of tone class B31, final vowel class *-e,* and stem-formative class 11 (see Table 4).

---

[20] In addition to the usefulness of single forms for probabilistically predicting other forms, it is sometimes the case that several forms within a paradigm more reliably predict forms than a single form alone (Paunonen 1976, Thymé 1993, Thymé et al. 1994) and sometimes the case that there are confederations of forms that partition a single paradigm into subparadigms of form predictability (Ackerman et al. 2009).

	NEUTRAL		INCOMPL			NEUTRAL		INCOMPL			NEUTRAL		INCOMPL	
	SG	PL	SG	PL		SG	PL	SG	PL		SG	PL	SG	PL
1INCL		2-2		4-42	1INCL		-ẽ		-ẽ	1INCL		ča-		ča-
1	3-1	2-24	3-1	4-24	1	-æ	-ĩ	-æ	-ĩ	1	ba-	ča-	kua-	ča-
2	2-2	2-2	4-2	4-2	2	-e	-ũ	-e	-ũ	2	ča-	ča-	ča-	ča-
3	3-2		4-2		3	-e		-e		3	ba-		kua-	

TABLE 4. Tone class B31, final vowel class *-e*, and stem-formative class 11 in Chiquihuitlán Mazatec.

Verbs' lexemes and word forms are distinguished by this network of paradigmatic relationships and alternations, and each of these separate inflectional systems by itself shows considerable complexity. Treated separately, their entropies (in Table 5) are individually in line with the paradigm entropies of inflection class systems discussed in the previous section.

	CELLS	REALIZATIONS	MAX REALIZATIONS	DECL CLASSES	DECL ENTROPY	PARADIGM ENTROPY	AVG ENTROPY
NEUTRAL TONES	6	16	4	6	2.585	0.264	1.622
FINAL VOWEL	6	11	9	10	3.322	0.775	1.333
STEM FORMATIVE	4	32	16	18	4.170	0.099	2.369

TABLE 5. Paradigm entropies for the Chiquihuitlán Mazatec inflectional systems.

In addition, each lexical item is a member of a META-CLASS, a combination of an inflection class from each of these three systems, and there are potentially 6 × 10 × 18 = 1,080 meta-classes. Baerman and Corbett (2010) report that 109 are attested in Capen's (1996) dictionary, most of which have only one or two members (the largest class has twenty-two members). This is a tiny fraction of the space of possible meta-classes. This fact does not help to reduce the overall complexity of the system, however, since for any particular lexeme, knowing the class membership in one dimension does little to help predict the class membership in another. On average, the uncertainty in guessing the class membership in one dimension is 2.469 bits, while the uncertainty in guessing the class membership in one dimension while knowing the class membership in another is 2.154 bits. So, the classes provide (on average) only 0.315 bits of information about each other. From this perspective CqM appears to be a worst case for inflection class complexity. Each lexeme is essentially in its own meta-class, where each meta-class is a conjunction of classes in three independently varying formal dimensions.

How does a system like CqM's look from the perspective of Carstairs-McCarthy's (2010) synonymy avoidance principle? Carstairs-McCarthy argues that purely affixal morphology is the syntagmatic combination of meaningful elements and so is expected to exhibit 'vocabular clarity'. In the context of inflection class inventories, this means that each realization of a paradigm cell should uniquely identify the inflection class membership of any stem it attaches to. In our terms, this means that paradigms are predicted to be organized in terms of mutual implication between all related word forms: average conditional entropy, as a consequence, should be zero. This can occur in the languages of the world, as it does in Northern Saami first declension nouns (Ackerman et al. 2009). However, the system of final vowels, given in Table 6, seems to dramatically violate expectations based on this principle: only the third-person forms are in any way diagnostic of class membership.

Furthermore, final vowel class membership does not apparently correlate with gender or any other syntactic or semantic class. According to Jamieson (1982), these patterns developed historically via the fusion of a stem-final vowel with a person/number

1SG	2SG	3	1INCL	1PL	2PL
-æ	-i	-i	-ẽ	-ĩ	-ũ
-æ	-e	-e	-ẽ	-ĩ	-ũ
-æ	-e	-æ	-ẽ	-ĩ	-ũ
-u	-i	-u	-ũ	-ĩ	-ũ
-o	-e	-o	-õ	-ĩ	-ũ
-a	-e	-a	-ã	-ĩ	-ũ
-ẽ	-ĩ	-ĩ	-ẽ	-ĩ	-ũ
-ẽ	-ĩ	-ẽ	-ẽ	-ĩ	-ũ
-ũ	-ĩ	-ũ	-ũ	-ĩ	-ũ
-ã	-ĩ	-ẽ	-ã	-ĩ	-ũ

Table 6. Chiquihuitlán Mazatec final vowel classes (Jamieson 1982:141).

suffix, so synchronically the final vowel class membership of a verb reflects only what that verb stem's final vowel was in an earlier stage of the language.

Turning to the stem formatives and the tone patterns, Carstairs-McCarthy (2010) argues that these kinds of stem alternations, unlike purely affixal morphology, should be seen in paradigmatic terms. Synonymy avoidance in this case is claimed to be reflected in 'distributional uniformity'. If, for example, some verbs in a language have one stem in the first- and third-person singular and a second stem in other forms, then synonymy avoidance would predict that all verbs with two stem alternants will distribute them in the same way. The system of stem formatives in CqM does appear to show this kind of distributional uniformity. For the most part, intransitive verbs have a single stem formative in the neutral aspect and transitive verbs have two. And, when a verb has two neutral aspect stem formatives, the first will be used for the third and first plural forms and the second will be used for other person and numbers.

Returning to CqM, we consider the tone patterns, given in Table 7.

1SG	2SG	3	1INCL	1PL	2PL
3-1	3-1	3-1	3-31	3-14	3-1
1-1	2-2	2-2	2-2	2-24	2-2
3-1	2-2	3-2	2-2	2-24	2-2
1-43	1-43	3-24	14-42	14-34	14-3
1-1	3-2	3-2	3-2	3-24	3-2
3-1	3-2	3-2	3-2	3-24	3-2

Table 7. Chiquihuitlán Mazatec neutral aspect tone patterns (Jamieson 1982:148).

Depending on one's point of view, the tone patterns could be seen as either affixation or as a stem alternation. However, from neither perspective are the predictions of synonymy avoidance borne out. None of the cells are diagnostic of class membership, so these tone patterns do not exhibit vocabular clarity. Nor do these tone patterns show distributional uniformity. For example, the first two classes have three distinct tone patterns, but in the first class it is the 1INCL and 1PL forms that are different, and in the second class it is the 1SG and 1PL.

In summary, Carstairs-McCarthy (2010) hypothesizes that affixes and stems reflect synonymy avoidance in distinct ways: the former syntagmatically and the latter paradigmatically. In both instances, actual morphological systems demonstrably deviate from the predictions of synonymy avoidance. In effect, synonymy avoidance identifies a state that morphological systems would occupy, all other things being equal. Naturally, one would like to know to what degree such departures can exist as well as what constrains the shape of these departures. Synonymy avoidance cannot account for such phenomena, nor can any single factor identified by Carstairs-McCarthy. We have suggested an answer to these questions that subsumes the effects of the synonymy avoidance principle and treats its syntagmatic and paradigmatic applications uniformly. The hypothesized effects of this principle are consistent with the low entropy conjecture, since in both dimensions the conditional entropy would be zero. Moreover, this conjecture predicts that deviations, whether syntagmatic or paradigmatic, should display low entropy. As a consequence, a single measure can be seen to comprehend a broad array of morphological organizations. It is also worth mentioning that the cognitive grounding of the competing sources of explanation differ. On Carstairs-McCarthy's account, synonymy avoidance itself is a general cognitive principle responsible for the idealized effects. In contrast, the low entropy conjecture suggests that cognitive principles concerning pattern organization are responsible for the development of learnable and usable morphological systems: on this analysis synonymy avoidance addresses particularly simple effects associated with low entropy.

As this brief outline of CqM's morphological structure has shown, CqM inflection is highly complex. Based on comparisons with nearby languages and reconstructed protolanguages, Jamieson (1982) offers diachronic explanations for the development of this surprising degree of complexity, but does not raise the question of how it might have been maintained. How are CqM speakers able to manage this complexity? Or, to put it another way, why has this system not been reorganized into a more coherent set of inflectional classes?

An answer to this latter question may be found by looking at the average conditional entropy of these verb paradigms. Guessing a word form of a verb requires choosing among the available tone patterns, final vowels, and stem formatives, each of which can vary independently. The average entropy for this problem (for positive neutral forms) is 4.920, equivalent to a choice among about thirty equally likely alternatives. However, the average conditional entropy—the uncertainty in guessing one word form given another—is only 0.709 bits, within the range of the sample of languages in Table 3.

There are at least three factors that contribute to CqM's relatively low I-complexity despite its high degree of E-complexity. One is that every word form provides speakers with information about all three inflectional dimensions simultaneously. The unusual organization of CqM's verbal inflection increases the uncertainty in each unknown word form but also increases the information provided by each known word form, leading to a system with an overall complexity similar to that of languages of more straightforward morphological types.

Another property of CqM that reduces the complexity of the system is that there is a significant degree of syncretism, both within and across paradigms. While there are a large number of distinct inflection classes, many of the classes differ in only one or two forms. For example, compare the paradigms in 19 above and 20 below.

(20) Positive paradigm for *ba^{3}šæ^2* 'take out' (Jamieson 1982:167)

	NEUTRAL		INCOMPLETIVE	
	SG	PL	SG	PL
1INCL		nã^{2}šẽ^2		nã^{2}šẽ^{42}
1	ba^{3}šæ^1	nã^{2}šĩ^{24}	kua^{3}šæ^1	nã^{4}šĩ^{34}
2	nã^{2}še^2	nã^{2}šũ^2	nã^{4}še^3	nã^{4}sũ^3
3	ba^{3}šæ^2		kua^{4}šæ^{41}	

The tone patterns, final vowels, and stem formatives used to express each of the cells in the paradigm in 20 differ from those in 19 in only a few ways. Listing these verbs as members of distinct inflectional classes greatly overstates the variation across lexemes and exaggerates the difficulty of predicting individual word forms.

The third and most important property of CqM that limits its average conditional entropy is the fact that its word forms are organized by a network of implicational relations that license reliable inferences about word forms. These relations are particularly important in a language with this level of E-complexity, as CqM does not achieve low average conditional entropy by being morphologically simple or by following strategies like the paradigm economy or no blur principles.

To test the role that implicational relations play in CqM morphology, we performed a simple 'bootstrap' simulation (Davison & Hinkley 1997). Statistical hypothesis testing proceeds by identifying a statistic whose sampling distribution is known under the null hypothesis H_0, and then estimating the probability of finding a result that deviates from what would be expected under H_0 at least as much as the observed data does. In this case, H_0 is that implicational relations are not a factor in reducing average conditional entropy in CqM, and the relevant statistic is the average conditional entropy. Unfortunately, we have no theoretical basis for deriving the sampling distribution of average conditional entropy under H_0, which precludes the use of conventional statistical methods. However, we can use a simple computational procedure for estimating the sampling distribution of the average conditional entropy. Take CqM′, an alternate version of CqM with the same E-complexity but with formal realizations assigned randomly to paradigm cells. More specifically, we generate CqM′ by constructing 109 random declensions, where each declension is produced by randomly selecting for each of the six paradigm cells one of the possible realizations of that cell. The result is a language with more or less the same E-complexity—the same number of declensions, paradigm cells, and allomorphs—as CqM, but with no implicational structure. Under the null hypothesis, implicational relations among paradigm cells are not crucial to reducing average conditional entropy, and so CqM′ should have more or less the same average conditional entropy as CqM. Any particular randomly generated CqM′ might have slightly higher or slightly lower average conditional entropy, so we repeat the comparison with many simulated CqM′s. The distribution of paradigm entropies across these randomly generated CqM′s provides an estimate of the sampling distribution of the average conditional entropy under H_0.

The results of this simulation are shown in Figure 2. No randomized CqM′ had an average conditional entropy as low as the actual language, and the average average conditional entropy for the randomized languages is 1.1 bits. The observed average conditional entropy is well outside what would be expected under the null hypothesis, which we can confidently reject in favor of the alternative, namely that assignment of realizations to cells in CqM is not in fact random. Instead, realizations in CqM are assigned to cells in a way that reduces average conditional entropy by licensing inferences about unknown word forms and is crucial to lowering its I-complexity.

In contrast, when we perform this same experiment with Russian, a language with relatively low E-complexity, we get a very different result: the average average conditional entropy of randomized versions of Russian is 0.541 bits, only slightly higher than the actual average conditional entropy of 0.538 bits. This indicates that the implicational structure of the Russian paradigm is much less important for constraining the overall average conditional entropy. Nearly any random mapping between the morphosyntactic property sets and the resources for exponence in Russian yields low entropy, so there is no need for such languages to rely on implicational organization.

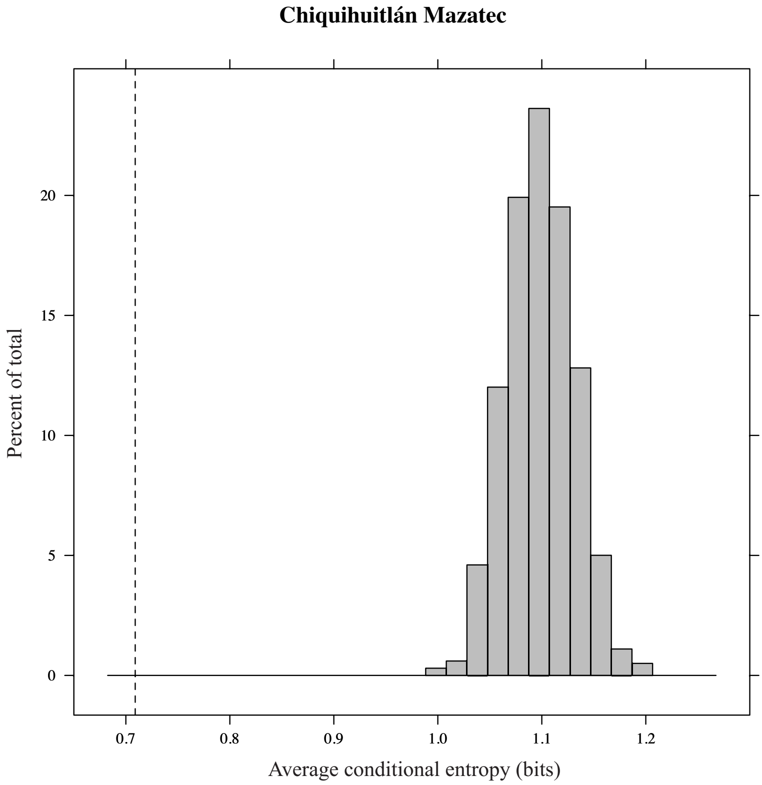

FIGURE 2. Distribution of average conditional entropies of 999 randomized versions of Chiquihuitlán Mazatec; the dashed line indicates the actual average conditional entropy.

There is another aspect of these simulation results that is important to note. All of the entropy calculations we make in this article are dependent on a range of basic assumptions. As noted, we look only at formal relationships. Any additional aspects of inflectional category membership that are predictable on the basis of phonological or semantic generalizations may lower the entropy further. Beyond that, the particular values we get for average conditional entropy depend on specific choices made in identifying paradigms for analysis and on enumerating paradigm cells and realizations. Different choices may well lead to slightly different values for paradigm entropies in the languages of our sample. What these simulations show is that the low paradigm entropies we have observed are not simply a by-product of our assumptions, since the high average conditional entropy of the simulated CqM′s follows from those same assumptions. Even given those assumptions, the average conditional entropies we have found are surprisingly low.

What we have tried to show in this section is that the complexity of the CqM verbal inflectional system is not very different from that of languages that appear on the surface to be much simpler. This is not to say that CqM is not a morphologically complex language, or that it is not in any way more complex than Russian or Modern Greek. Writing a description of CqM inflection classes or compiling a CqM dictionary would be dramatically more difficult than doing so for Modern Greek, as each individual word form carries more uncertainty. This, however, is a measure of CqM’s E-complexity and does not reflect the language's I-complexity, which reflects the difficulty the language poses for speakers who are attempting to use CqM to communicate. From this perspective, CqM is not especially complex, the inflectional system poses no impediment to using the language, and there would be no pressure for the system of classes to reorganize or simplify.

5. CONCLUSIONS. In this article we have built upon word-and-paradigm morphology, a tradition of morphological analysis that existed prior to various variants of structuralist and generative morphology and has developed in parallel to it over the past fifty years.[21] In this tradition words are generally posited to be the smallest units of morphological analysis. This permits them to enter into calculations of relatedness among word forms, both relatedness among forms of a specific lexeme and patterns of relatedness associated with classes of lexemes.[22] The organization of morphology along these lines encourages speculation about what facilitates the learnability of such systems and what constrains the organization of inflectional classes within a given language.

Concerning the former, there have been efforts to identify diagnostic forms that are useful for predicting unknown forms of known lexemes and, thereby, identifying inflection class membership for specific lexemes. This has yielded a rich tradition regarding principal parts (Stump & Finkel 2007, 2009), and their role can be construed more broadly in terms of implicative relations among members of paradigms (Wurzel 1989, Blevins 2006). But, the special diagnostic value of principal parts turns out, on the examination of complex morphological systems, to be a property shared to some degree by all forms in paradigms. By looking at the average conditional entropy of individual paradigm cells (as in Table 2), we can recognize predictiveness and predictability as gradient concepts that follow more generally from the organization of the paradigm.

With respect to the number of inflection classes, hypotheses such as the paradigm economy principle and the no blur principle have attempted to explain limits on their number on the basis of E-complexity properties of particular languages. Namely, the number of inflection classes is claimed to be contingent on the largest number of allomorphs associated with any cell in a paradigm: the number of inflectional classes is predicted not to exceed the number of allomorphs contained in such a cell. However, claims about the number of inflection classes often depart considerably from what the relevant principles predict. Carstairs-McCarthy (2010) proposes an explanatory constraint on E-complexity for the effects of the no blur principle in terms of vocabular clarity, a cognitive principle.

On the present account, we have proposed a measure, average conditional entropy, that quantifies what it means for forms to participate in implicative relations, recognizing that all forms have some diagnostic value. While our conditional entropy measures provide insight into the organization of static, synchronic paradigms, there are crucial complementary questions about how such patterns are learned. In fact, these results are

---

[21] See, for example, Robins 1959, Matthews 1974, Jackendoff 1975, Bochner 1993, Blevins 2006, 2013c.

[22] It is worth observing that an emphasis on whole words does not preclude the possibility that the construct 'morpheme' is useful in some languages in some instances. In other words, there seems empirical warrant to take a pragmatic view on the structure of words, avoiding reductive claims that all words must have morphemes or that whole words cannot have evidently morphemic components. As observed by Blevins (2013a), morphological organization with respect to morphemes in classic agglutinative languages is straightforwardly simple in terms of conditional entropy measures: in the simplest case, the selection of a single marker to express a single property has an entropy of zero, since the system ensures that there are no surprises in morphological expression.

consistent with research in which constraints on synchronic grammar are provided by constraints on usage, language development, and learnability in linguistic as well as nonlinguistic domains.[23] In this connection, the information-theoretic measures we have used and the strategy of appealing to computational simulations are the tools that have been increasingly employed within developmental sciences for the insightful analysis of patterns within complex adaptive systems in both biology and psychology. With respect to psychology in particular, there has been a recent reappraisal of the tutelary value of earlier literature concerning information-theoretic analyses of pattern perception and the relevance of this for development that derives from the 'discriminative learning' research in animal behavior (Miller 1953, Rescorla 1988a,b, Ramscar & Dye 2009, Ramscar et al. 2010). This latter literature, accordingly, offers explicit strategies and insights for exploring the results obtained here from synchronic analysis to their possible developmental pathways.

We believe that appealing to information-theoretic measures helps to detect patterns and generalizations that exist otherwise undetected in the data and, therefore, constitute reliable bases for theoretical speculation. They permit us to explore and make sense of variation itself as a signally instructive source concerning the nature of morphological systems. As mentioned, the basic analytic objects of our inquiry have been whole words and their organization into networks or patterns of relatedness. It is by regarding words and paradigms as primary objects of linguistic analysis that we have conjectured a fundamental organizing principle of morphological organization, namely, low conditional entropy among (patterns of) words. This therefore counts as evidence against arguments dismissing words and paradigms as epiphenomena (a view typified by Julien 2002 and Embick & Marantz 2008). Correlatively, the very existence of reliably low entropies in complex morphologies implicates the irreducible effects of patterns in grammar.

In particular, since low entropies reflect patterned organization, grammatical approaches that posit words and paradigms and clausal constructions as primary units of analysis appear to be well grounded. If low entropy is the correct measure for explaining the implicational organization of paradigms, rendering complex systems learnable, then this makes a prediction about types of systems that we do not find. While there are generally pockets of suppletion in any morphological system, that is, opaque and unpredictable forms of related words, there are no known fully suppletive systems in the languages of the world. This absence is easily explicable given the low entropy conjecture and its facilitating function for learnability: a completely suppletive system is one in which no form bears an implicational relation with any other form, and thus there is no useful patterned organization reflective of low entropy. Instead, such a system, precluding the possibility of generalizing from direct experience to future forms, would appear to demand a colossal memory task. Though we are obviously capable of storing huge quantities of individual words, it appears that inflectional morphology intelligently organizes otherwise random mappings into more or less transparent patterns of relatedness.

In sum, inflectional systems may be able to tolerate untold degrees of E-complexity as long as I-complexity, interpreted as low average conditional entropy, resides in the shape of its patterned organization (though of course we do not preclude the existence of additional conditioning factors). It appears that information-theoretic analysis is a significant resource for providing insights into the structure and organization of crosslinguistic morphological systems when words and paradigms are interpreted as primary objects of analysis.

---

[23] See, for example, Karmiloff-Smith 1994, 1998, Bates 1998, 1999, Thelen & Bates 2003, Tomasello 2003, Bybee 2006, Bannard & Matthews 2008, among others. Of particular relevance is the literature on distributional learning of language patterns typified in Braine 1966, 1987, Mintz 2002, Gerken 2009, Gervain & Erras 2012, Gervain & Werker 2013.

Appendix: Paradigms for language sample

CLASS	1SG	2SG	3SG
1	-ni	-n	-g
4	-ni	-n	-ug
5	-eni	-ein	-ug
6	-ni	-n	-nag
7	-ni	-in	-nug
8	-mi	-m	-g
9	-ni	-n	-n
11	-ni	-n	-c
13	-ni	-in	-c
14	-ni	-in	-ic
16	-mi	-im	-c
17	-mi	-m	-h
18	-mi	-im	-h
19	-ni	-n	-Ø
22	-ini	-inin	-Ø
23	-ani	-ain	-Ø
25	-ni	-in	-ig
26	-li	-in	-ig
27	-i	-in	-ig
28	-i	-im	-ig
29	-i	-in	-iag
30	-i	-in	-ag
31	-i	-en	-eg

Table A1. Amele possessor suffixes (Roberts 1987, Hein & Müller 2009).

CLASS	SG	PL
1	-b^{y}	-bys
2	-bŕr	-ryb
3	-g	-(ga)s
4	-k^{u}	-meb
5	-k^{u}	-u
6	-k^{u}	-rib
7	-k^{u}	-ib
8	-k^{u}	-guhijer
9	-k^{u}	-ijer
10	-k^{u}	-komi
11	-k^{u}	-heu
12	-Vm	-(e)ip^{i}
13	-n	-b
14	-n	-ab
15	-n	-M
16	-ñ	-š
17	-V	-has
18	-p^{u}	-gwis
19	-p^{u}	-s
20	-r	-gu^{h}
21	-t	-g^{u}
22	-t	-tog^{u}
23	-uh^{uh}	-ruh
24	-^{uh}	-gwiruh
25	-a^{h}	-e^{h}
26	-u^{h}	-i^{h}

Table A2. Arapesh number suffixes (Dobrin 1999, Hein & Müller 2009).

CLASS	I.SG	I.PL	II.SG	II.PL	III.SG	III.PL	IV.SG	IV.PL	V.SG	V.PL	VI.SG	VI.PL
A	j-	s-	g-	s-	g-	j-	j-	j-	j-	g-	g-	g-
B	b-	t-	n-	t-	n-	b-	b-	b-	b-	n-	n-	n-

Table A3. Burmeso object agreement prefixes (Donohue 2001, Baerman et al. 2010).

	1/2 SG/PL			3 PL [+HUM]			3 PL [−HUM]			3 SG		
CLASS	PERF	PRES	SUBJ	PERF	PRES	SUBJ	PERF	PRES	SUBJ	PERF	PRES	SUBJ
I,1a	LH-ò	LH-èl	LH-o	LH-ùl	LH-èl-à/ì	LH-òl	HH-ùl	HH-èl-à/ì	HH-òl	HH-ò	HH-èl	HH-o
I,1b	LH-ò	LF-Ø	LH-o	LH-ùl	LH-è	LH-òl	HH-ùl	HH-è	HH-òl	HH-ò	HF-Ø	HH-o
I,1c	LH-ò	LH-ì	LH-o	LH-ùl	LH-è	LH-òl	HH-ùl	HH-è	HH-òl	HH-ò	HH-ì	HH-o
I,2a	HH-o	HH-èl	HH-ò	HH-ùl	HH-èl-à/ì	HH-òl	LL-ùl	LL-èl-à/ì	LL-òl	LL-ò	LL-èl	LL-o
I,2b	HH-o	HF-Ø	HH-ò	HH-ùl	HH-è	HH-òl	LL-ùl	LL-è	LL-òl	LL-ò	LL-Ø	LL-o
I,2c	HH-o	HH-ì	HH-ò	HH-ùl	HH-è	HH-òl	LL-ùl	LL-è	LL-òl	LL-ò	LL-ì	LL-o
II,1a	LH-i	LH-itì	LH-i	LH-i-è	LH-iti-A(l)	LH-i-A(l)	HH-i-è	HH-iti-A(l)	HH-i-A(l)	HH-i	HH-itì	HH-i
II,1b	LH-i	LF-Ø	LH-i	LH-i-è	LH-è	LH-i-A(l)	HH-i-è	HH-è	HH-i-A(l)	HH-i	HF-Ø	HH-i
II,2a	HH-ì	HH-itì	HH-ì	HH-i-è	HH-iti-A(l)	HH-i-A(l)	LL-i-è	LL-iti-A(l)	LL-i-A(l)	LL-i	LL-itì	LL-i
II,2b	HH-ì	HF-Ø	HH-ì	HH-i-è	HH-è	HH-i-A(l)	LL-i-è	LL-è	LL-i-A(l)	LL-i	LF-Ø	LL-i
IIIa	HH-à	HH-èl	HH-ì	HH-e	HH-èl-à	HH-è	LH-e	LH-èl-à	LH-è	LH-à	LH-èl	LH-ì
IIIb	HH-ò	HH-èl	HH-ò	HH-e	HH-èl-à	HH-è	LH-e	LH-èl-à	LH-è	LH-ò	LH-èl	LH-ò
IIIc	HH-ò	HH-èl	HH-ò	HH-e	HH-èl-à	HH-è	LH-e	LH-èl-à	LH-è	LH-ò	LH-èl	LF-Ø
IIId	HH-à	HH-èl	HF-Ø	HH-e	HH-èl-à	HH-è	LH-e	LH-èl-à	LH-è	LH-à	LH-èl	LF-Ø
IIIe	HH-à	HH-èl	HF-Ø	HH-e	HH-èl-à	HH-è	LH-e	LH-èl-à	LH-è	LH-ò	LH-èl	LF-Ø
IVa	HH-ò	HH-èl	HF-Ø	HH-e	HH-èl-à	HH-Al	LH-e	LH-èl-à	LH-Al	LH-ò	LH-èl	LF-Ø
IVb	HH-ò	HH-èl	HH-ò	HH-e	HH-èl-à	HH-Al	LH-e	LH-èl-à	LH-Al	LH-ò	LH-èl	LH-ò
IVc	HH-à	HH-èl	HF-Ø	HH-e	HH-èl-à	HH-Al	LH-e	LH-èl-à	LH-Al	LH-à	LH-èl	LF-Ø
IVd	HH-à	HH-èl	HH-à	HH-e	HH-èl-à	HH-Al	LH-e	LH-èl-à	LH-Al	LH-à	LH-èl	LH-à

Table A4. Fur verbal tense and agreement markers (tones and suffixes) (Jakobi 1990, Stump & Finkel 2007).

	1				2				3			
CLASS	SG	SG.DIM	DU	PL	SG	SG.DIM	DU	PL	SG	SG.DIM	DU	PL
I	a	a	ac	ec	a	a	ac	ac	a	a	ac	naN
II	a	naN	aN	eN	a	naN	aN	aN	a	naN	aN	naN
III	a	naN	aN	e	a	naN	aN	a	a	naN	aN	a
IV	a	naN	aN	era	a	naN	aN	ara	a	naN	aN	ara

Table A5. Kwerba subject agreement prefixes (De Vries & De Vries 1997, Stump & Finkel 2007).

CLASS	1SG	2SG	3	1INCL	1PL	2PL
1	a/1-1/tsi	e/2-2/nĩ	a/(2-2)/tsi	ã/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
2	a/3-1/tsi	e/3-1/nĩ	a/(3-1)/tsi	ã/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
3	æ/1-1/tsi	i/2-2/nĩ	i/(2-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
4	u/1-1/tsi	i/2-2/nĩ	u/(2-2)/tsi	ũ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
5	a/1-1/bi	e/2-2/bi	a/(2-2)/bi	ã/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
6	a/3-1/bi	e/2-2/bi	a/(3-2)/bi	ã/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
7	u/3-1/tsi	i/3-1/nĩ	u/(3-1)/tsi	ũ/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
8	æ/1-1/tsi	e/2-2/nĩ	e/(2-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
9	o/3-1/tsi	e/2-2/nĩ	o/(3-2)/tsi	õ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
10	ẽ/1-43/bi	ĩ/1-43/bi	ẽ/(3-24)/bi	ẽ/14-42/bi	ĩ/14-34/bi	ũ/14-3/bi
11	æ/3-1/tsi	i/3-1/nĩ	i/(3-1)/tsi	ẽ/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
12	æ/1-1/tsi	e/2-2/nĩ	æ/(2-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
13	a/1-43/tsi	e/1-43/nĩ	a/(3-24)/tsi	ã/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
14	ẽ/1-43/tsi	ĩ/1-43/nĩ	ẽ/(3-24)/tsi	ẽ/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
15	u/1-43/tsi	i/1-43/nĩ	u/(3-24)/tsi	ũ/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
16	ẽ/3-1/bu	ĩ/3-1/ču	ẽ/(3-1)/bu	ẽ/3-31/ču	ĩ/3-14/ču	ũ/3-1/ču
17	æ/3-1/tsi	e/3-1/nĩ	æ/(3-1)/tsi	ẽ/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
18	ũ/1-1/tsi	ĩ/2-2/nĩ	ũ/(2-2)/tsi	ũ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
19	æ/3-1/bi	e/3-1/bi	e/(3-1)/bi	ẽ/3-31/bi	ĩ/3-14/bi	ũ/3-1/bi
20	a/1-1/be	e/2-2/be	a/(2-2)/be	ã/2-2/be	ĩ/2-24/be	ũ/2-2/be
21	u/1-43/be	i/1-43/be	u/(3-24)/be	ũ/14-42/be	ĩ/14-34/be	ũ/14-3/be
22	a/1-1/ba	e/2-2/ba	a/(2-2)/ba	ã/2-2/ba	ĩ/2-24/ba	ũ/2-2/ba
23	a/3-1/ba	e/2-2/ba	a/(3-2)/ba	ã/2-2/ba	ĩ/2-24/ba	ũ/2-2/ba
24	a/1-1/bu	e/2-2/ču	a/(2-2)/bu	ã/2-2/ču	ĩ/2-24/ču	ũ/2-2/ču
25	ẽ/1-43/bu	ĩ/1-43/ču	ẽ/(3-24)/bu	ẽ/14-42/ču	ĩ/14-34/ču	ũ/14-3/ču
26	a/3-1/hba	e/2-2/hba	a/(3-2)/hba	ã/2-2/hba	ĩ/2-24/hba	ũ/2-2/hba
27	æ/3-1/hba	i/2-2/hba	i/(3-2)/hba	ẽ/2-2/hba	ĩ/2-24/hba	ũ/2-2/hba
28	ẽ/1-1/tsi	ĩ/2-2/nĩ	ẽ/(2-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
29	æ/3-1/tsi	e/2-2/nĩ	æ/(3-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
30	ẽ/3-1/tsi	ĩ/2-2/nĩ	ĩ/(3-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
31	æ/3-1/tsi	i/3-1/nĩ	i/(3-1)/tsi	ẽ/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
32	æ/1-43/tsi	i/1-43/nĩ	i/(3-24)/tsi	ẽ/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
33	æ/1-1/tsi	e/3-2/nĩ	æ/(3-2)/tsi	ẽ/3-2/nĩ	ĩ/3-24/nĩ	ũ/3-2/nĩ
34	a/3-1/ba	e/3-1/ča	a/(3-1)/ba	ã/3-31/ča	ĩ/3-14/ča	ũ/3-1/ča
35	æ/3-1/ba	e/2-2/ča	e/(3-2)/ba	ẽ/2-2/ča	ĩ/2-24/ča	ũ/2-2/ča
36	æ/1-1/bi	i/2-2/bi	i/(2-2)/bi	ẽ/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
37	ũ/1-1/bi	ĩ/2-2/bi	ũ/(2-2)/bi	ũ/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
38	u/3-1/bi	i/2-2/bi	u/(3-2)/bi	ũ/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
39	a/1-43/bi	e/1-43/bi	a/(3-24)/bi	ã/14-42/bi	ĩ/14-34/bi	ũ/14-3/bi
40	a/1-1/bu	e/2-2/ntu	a/(2-2)/bu	ã/2-2/ntu	ĩ/2-24/ntu	ũ/2-2/ntu
41	ẽ/1-43/bu	ĩ/1-43/ntu	ẽ/(3-24)/bu	ẽ/14-42/ntu	ĩ/14-34/ntu	ũ/14-3/ntu
42	a/3-1/be	e/3-1/be	a/(3-1)/be	ã/3-31/be	ĩ/3-14/be	ũ/3-1/be
43	æ/3-1/be	i/3-1/be	i/(3-1)/be	ẽ/3-31/be	ĩ/3-14/be	ũ/3-1/be
44	o/3-1/be	e/3-1/be	o/(3-1)/be	õ/3-31/be	ĩ/3-14/be	ũ/3-1/be
45	ũ/3-1/be	ĩ/3-1/be	ũ/(3-1)/be	ũ/3-31/be	ĩ/3-14/be	ũ/3-1/be
46	æ/1-1/be	i/2-2/be	i/(2-2)/be	ẽ/2-2/be	ĩ/2-24/be	ũ/2-2/be
47	o/1-1/be	e/2-2/be	o/(2-2)/be	õ/2-2/be	ĩ/2-24/be	ũ/2-2/be
48	a/3-1/be	e/2-2/be	a/(3-2)/be	ã/2-2/be	ĩ/2-24/be	ũ/2-2/be
49	ã/3-1/be	ĩ/2-2/be	ẽ/(3-2)/be	ã/2-2/be	ĩ/2-24/be	ũ/2-2/be

(Table A6. *Continues*)

CLASS	1SG	2SG	3	1INCL	1PL	2PL
50	æ/3-1/be	i/2-2/be	i/(3-2)/be	ẽ/2-2/be	ĩ/2-24/be	ũ/2-2/be
51	u/3-1/be	i/2-2/be	u/(3-2)/be	ũ/2-2/be	ĩ/2-24/be	ũ/2-2/be
52	a/1-43/be	e/1-43/be	a/(3-24)/be	ã/14-42/be	ĩ/14-34/be	ũ/14-3/be
53	æ/1-43/be	i/1-43/be	i/(3-24)/be	ẽ/14-42/be	ĩ/14-34/be	ũ/14-3/be
54	ẽ/1-43/be	ĩ/1-43/be	ĩ/(3-24)/be	ẽ/14-42/be	ĩ/14-34/be	ũ/14-3/be
55	æ/3-1/ba	e/3-1/ba	e/(3-1)/ba	ẽ/3-31/ba	ĩ/3-14/ba	ũ/3-1/ba
56	ũ/3-1/ba	ĩ/3-1/ba	ũ/(3-1)/ba	ũ/3-31/ba	ĩ/3-14/ba	ũ/3-1/ba
57	æ/3-1/ba	e/3-1/ba	æ/(3-1)/ba	ẽ/3-31/ba	ĩ/3-14/ba	ũ/3-1/ba
58	æ/1-1/ba	e/2-2/ba	æ/(2-2)/ba	ẽ/2-2/ba	ĩ/2-24/ba	ũ/2-2/ba
59	æ/1-1/ba	i/2-2/ba	i/(2-2)/ba	ẽ/2-2/ba	ĩ/2-24/ba	ũ/2-2/ba
60	ã/3-1/ba	ĩ/2-2/ba	ẽ/(3-2)/ba	ã/2-2/ba	ĩ/2-24/ba	ũ/2-2/ba
61	ẽ/3-1/ba	ĩ/2-2/ba	ĩ/(3-2)/ba	ẽ/2-2/ba	ĩ/2-24/ba	ũ/2-2/ba
62	æ/1-43/ba	i/1-43/ba	i/(3-24)/ba	ẽ/14-42/ba	ĩ/14-34/ba	ũ/14-3/ba
63	u/1-43/ba	i/1-43/ba	u/(3-24)/ba	ũ/14-42/ba	ĩ/14-34/ba	ũ/14-3/ba
64	o/1-43/bo	e/1-43/čo	o/(3-24)/bo	õ/14-42/čo	ĩ/14-34/čo	ũ/14-3/čo
65	a/3-1/bu	e/3-1/ču	a/(3-1)/bu	ã/3-31/ču	ĩ/3-14/ču	ũ/3-1/ču
66	ẽ/3-1/bu	ĩ/2-2/ču	ẽ/(3-2)/bu	ẽ/2-2/ču	ĩ/2-24/ču	ũ/2-2/ču
67	a/1-43/bu	e/1-43/ču	a/(3-24)/bu	ã/14-42/ču	ĩ/14-34/ču	ũ/14-3/ču
68	a/1-1/hu	e/2-2/čhu	a/(2-2)/hu	ã/2-2/čhu	ĩ/2-24/čhu	ũ/2-2/čhu
69	æ/1-1/hu	e/2-2/čhu	æ/(2-2)/hu	ẽ/2-2/čhu	ĩ/2-24/čhu	ũ/2-2/čhu
70	a/3-1/hi	e/3-1/čhi	a/(3-1)/hi	ã/3-31/čhi	ĩ/3-14/čhi	ũ/3-1/čhi
71	ũ/3-1/hi	ĩ/3-1/čhi	ũ/(3-1)/hi	ũ/3-31/čhi	ĩ/3-14/čhi	ũ/3-1/čhi
72	a/1-1/hi	e/2-2/čhi	a/(2-2)/hi	ã/2-2/čhi	ĩ/2-24/čhi	ũ/2-2/čhi
73	æ/3-1/hba	i/3-1/hba	i/(3-1)/hba	ẽ/3-31/hba	ĩ/3-14/hba	ũ/3-1/hba
74	æ/1-43/hba	i/1-43/hba	i/(3-24)/hba	ẽ/14-42/hba	ĩ/14-34/hba	ũ/14-3/hba
75	ũ/3-1/tsi	ĩ/3-1/nĩ	ũ/(3-1)/tsi	ũ/3-31/nĩ	ĩ/3-14/nĩ	ũ/3-1/nĩ
76	ã/1-1/tsi	ĩ/2-2/nĩ	ẽ/(2-2)/tsi	ã/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
77	a/3-1/tsi	e/2-2/nĩ	a/(3-2)/tsi	ã/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
78	æ/3-1/tsi	e/2-2/nĩ	æ/(3-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
79	ẽ/3-1/tsi	ĩ/2-2/nĩ	ẽ/(3-2)/tsi	ẽ/2-2/nĩ	ĩ/2-24/nĩ	ũ/2-2/nĩ
80	æ/1-43/tsi	e/1-43/nĩ	æ/(3-24)/tsi	ẽ/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
81	ẽ/1-43/tsi	ĩ/1-43/nĩ	ĩ/(3-24)/tsi	ẽ/14-42/nĩ	ĩ/14-34/nĩ	ũ/14-3/nĩ
82	a/1-1/tsi	e/3-2/nĩ	a/(3-2)/tsi	ã/3-2/nĩ	ĩ/3-24/nĩ	ũ/3-2/nĩ
83	a/1-1/tsi	e/3-2/nĩ	a/(3-2)/tsi	ã/3-2/nĩ	ĩ/3-24/nĩ	ũ/3-2/nĩ
84	ã/3-1/tsi	ĩ/3-2/nĩ	ẽ/(3-2)/tsi	ã/3-2/nĩ	ĩ/3-24/nĩ	ũ/3-2/nĩ
85	æ/3-1/su	e/3-1/nũ	e/(3-1)/su	ẽ/3-31/nũ	ĩ/3-14/nũ	ũ/3-1/nũ
86	æ/3-1/ba	e/3-1/ča	æ/(3-1)/ba	ẽ/3-31/ča	ĩ/3-14/ča	ũ/3-1/ča
87	æ/3-1/ba	i/3-1/ča	i/(3-1)/ba	ẽ/3-31/ča	ĩ/3-14/ča	ũ/3-1/ča
88	æ/3-1/ka	i/3-1/ča	i/(3-1)/ka	ẽ/3-31/ča	ĩ/3-14/ča	ũ/3-1/ča
89	a/3-1/hba	e/3-1/nã	a/(3-1)/hba	ã/3-31/nã	ĩ/3-14/nã	ũ/3-1/nã
90	æ/3-1/ba	e/2-2/nã	æ/(3-2)/ba	ẽ/2-2/nã	ĩ/2-24/nã	ũ/2-2/nã
91	a/3-1/bi	e/3-1/bi	a/(3-1)/bi	ã/3-31/bi	ĩ/3-14/bi	ũ/3-1/bi
92	ã/3-1/bi	ĩ/3-1/bi	ẽ/(3-1)/bi	ã/3-31/bi	ĩ/3-14/bi	ũ/3-1/bi
93	u/3-1/bi	i/3-1/bi	u/(3-1)/bi	ũ/3-31/bi	ĩ/3-14/bi	ũ/3-1/bi
94	ã/1-1/bi	ĩ/2-2/bi	ẽ/(2-2)/bi	ã/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
95	æ/1-1/bi	e/2-2/bi	e/(2-2)/bi	ẽ/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
96	æ/1-1/bi	e/2-2/bi	æ/(2-2)/bi	ẽ/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
97	ã/3-1/bi	ĩ/2-2/bi	ẽ/(3-2)/bi	ã/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
98	æ/3-1/bi	i/2-2/bi	i/(3-2)/bi	ẽ/2-2/bi	ĩ/2-24/bi	ũ/2-2/bi
99	æ/1-43/bi	i/1-43/bi	i/(3-24)/bi	ẽ/14-42/bi	ĩ/14-34/bi	ũ/14-3/bi
100	a/1-1/bi	e/3-2/bi	a/(3-2)/bi	ã/3-2/bi	ĩ/3-24/bi	ũ/3-2/bi
101	a/3-1/bi	e/3-2/bi	a/(3-2)/bi	ã/3-2/bi	ĩ/3-24/bi	ũ/3-2/bi
102	a/3-1/bi	e/3-2/bi	a/(3-2)/bi	ã/3-2/bi	ĩ/3-24/bi	ũ/3-2/bi
103	u/1-1/bu	i/2-2/ntu	u/(2-2)/bu	ũ/2-2/ntu	ĩ/2-24/ntu	ũ/2-2/ntu
104	a/1-1/bu	e/3-2/ntu	a/(3-2)/bu	ã/3-2/ntu	ĩ/3-24/ntu	ũ/3-2/ntu
105	a/3-1/hba	e/3-1/čha	a/(3-1)/hba	ã/3-31/čha	ĩ/3-14/čha	ũ/3-1/čha
106	æ/3-1/hba	e/3-1/čha	e/(3-1)/hba	ẽ/3-31/čha	ĩ/3-14/čha	ũ/3-1/čha

(TABLE A6. *Continues*)

CLASS	1SG	2SG	3	1INCL	1PL	2PL
107	ã/3-1/hba	ĩ/3-1/čha	ẽ/(3-1)/hba	ã/3-31/čha	ĩ/3-14/čha	ũ/3-1/čha
108	a/1-1/hba	e/2-2/čha	a/(2-2)/hba	ã/2-2/čha	ĩ/2-24/čha	ũ/2-2/čha
109	a/3-1/hba	e/2-2/čha	a/(3-2)/hba	ã/2-2/čha	ĩ/2-24/čha	ũ/2-2/čha

Table A6. Mazatec verb tense and agreement markers (Jamieson 1982, Capen 1996, Baerman & Corbett 2010).

CLASS	IMP PL	IMP SG	IMPERF DIST FUT	IMPERF NEAR FUT	IMPERF PAST COND	IMPERF PAST CONT	IMPERF PAST HAB	INF
v1a	a-L	a-L	a-L	a-L	a-L	a-L	a-H	a-L
v1b	o-L	o-L	o-H	o-H	o-L	o-L	o-H	o-L
v2a.itr	a-M	a-M	a-M	a-M	a-L	a-M	a-H	a-M
v2a.tr	a-M	a-M	a-M	a-M	a-L	a-M	a-H	a-M
v2b.itr	o-M	o-M	o-M	o-M	o-L	o-L	o-H	o-M
v2b.tr	o-M	o-M	o-M	o-M	o-L	o-L	o-H	o-M
v3.itr	a-M	a-LM	a-L	a-L	a-L	a-L	a-H	a-M
v3.tr	a-M	a-LM	a-M	a-M	a-L	a-M	a-H	a-M
v4.itr	a-LM	a-LM	a-H	a-H	a-L	a-H	a-H	a-H
v4.tr	a-LM	a-LM	a-H	a-H	a-L	a-H	a-H	a-H

CLASS	NOM1	NOM2	PERF INDEF PAST	PERF NEAR PAST	PERF PRES	PERF REC PAST	PERF REM PAST	SUBJ
v1a	a-L	a-L	a-L	a-L	a-L	a-L	a-L	a-L
v1b	o-L	o-L	o-L	o-L	o-L	o-L	o-L	o-H
v2a.itr	a-M	a-M	a-L	a-M	a-L	a-L	a-L	a-M
v2a.tr	a-M	a-L	a-L	a-M	a-L	a-L	a-L	a-M
v2b.itr	o-M	o-M	o-L	o-L	o-L	o-L	o-L	o-M
v2b.tr	o-M	o-L	o-L	o-L	o-L	o-L	o-L	o-M
v3.itr	a-M	a-M	a-M	a-L	a-M	a-M	a-M	a-L
v3.tr	a-LM	a-M	a-M	a-M	a-M	a-M	a-M	a-M
v4.itr	a-LM	a-LM	a-LM	a-H	a-LM	a-LM	a-LM	a-H
v4.tr	a-LM	a-M	a-LM	a-H	a-LM	a-LM	a-LM	a-H

Table A7. Ngiti verbal inflection (stem-initial vowel and tone) (Kutsch Lojenga 1994, Stump & Finkel 2007).

CLASS	GEN.PL	GEN.SG	LOC.PL	LOC.SG	NOM.PL	NOM.SG
I.I	-Ø	-Ø	-ni̱	-kä	-Ø	-Ø
I.II	-ni̱	-Ø	-ni̱	-kä	-ni̱	-Ø
I.III	-ni̱	-Ø	-ni̱	-kä	-Ø	-Ø
II.I	-Ø	-kä	-ni̱	-kä	-Ø	-Ø
II.II	-ni̱	-kä	-ni̱	-kä	-ni̱	-Ø
II.III	-ni̱	-kä	-ni̱	-kä	-Ø	-Ø
II.IV	-ni̱	-kä	-Ø	-kä	-Ø	-Ø
II.V	-Ø	-kä	-ni̱	-kä	-ni̱	-Ø
II.VI	-Ø	-kä	-Ø	-kä	-Ø	-Ø
III.II	-ni̱	-kä	-ni̱	-Ø	-ni̱	-Ø
III.III	-ni̱	-kä	-ni̱	-Ø	-Ø	-Ø
III.IV	-ni̱	-kä	-Ø	-Ø	-Ø	-Ø
IV.II	-ni̱	-Ø	-ni̱	-Ø	-ni̱	-Ø
IV.III	-ni̱	-Ø	-ni̱	-Ø	-Ø	-Ø
IV.IV	-ni̱	-Ø	-Ø	-Ø	-Ø	-Ø
IV.V	-Ø	-Ø	-Ø	-Ø	-Ø	-Ø

Table A8. Nuer case suffixes (Frank 1999, Baerman et al. 2010).

	SINGULAR						PLURAL					
CLASS	NOM	ACC	GEN	DAT	LOC	INST	NOM	ACC	GEN	DAT	LOC	INST
I	-o	-o	-a	-u	-e	-om	-a	-a	-Ø	-am	-ax	-am'i
II	-Ø	-Ø	-a	-u	-e	-om	-i	-i	-ov	-am	-ax	-am'i
III	-a	-u	-i	-e	-e	-oj	-i	-i	-Ø	-am	-ax	-am'i
IV	-Ø	-Ø	-i	-i	-i	-ju	-i	-i	-ej	-am	-ax	-am'i

Table A9. Russian case suffixes (Baerman et al. 2010).

REFERENCES

Ackerman, Farrell; James P. Blevins; and Robert Malouf. 2009. Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In Blevins & Blevins, 54–82.

Ackerman, Farrell, and Gregory Stump. 2004. Paradigms and periphrasis: A study in realization-based lexicalism. *Projecting morphology*, ed. by Louisa Sadler and Andrew Spencer, 111–57. Stanford, CA: CSLI Publications.

Alberch, Pere. 1989. The logic of monsters: Evidence for internal constraint in development and evolution. *Geobios* 22.21–57.

Albright, Adam. 2005. The morphological basis of paradigm leveling. *Paradigms in phonological theory*, ed. by Laura Downing, Tracy Alan Hall, and Renate Raffelsiefen, 17–43. Oxford: Oxford University Press.

Albright, Adam. 2009. Modeling analogy as probabilistic grammar. In Blevins & Blevins, 185–213.

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90.119–61.

Anderson, Stephen R. 1985. *Phonology in the twentieth century: Theories of rules and theories of representations*. Chicago: University of Chicago Press.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Aronoff, Mark. 1993. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.

Baayen, R. Harald; Petar Milin; Dušica Filipović Đurđević; Peter Hendrix; and Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118.438–82.

Baerman, Matthew. 2012. Paradigmatic chaos in Nuer. *Language* 88.467–94.

Baerman, Matthew; Dunstan Brown; and Greville G. Corbett. 2010. Morphological complexity: A typological perspective. Surrey: University of Surrey, ms. Online: http://www.morphology.surrey.ac.uk/Papers/Morphological_complexity.pdf.

Baerman, Matthew, and Greville G. Corbett. 2010. A typology of inflectional class interaction. Paper presented at the 14th International Morphology Meeting, Budapest.

Bannard, Colin, and Danielle E. Matthews. 2008. Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19.241–48.

Bates, Elizabeth. 1998. Construction grammar and its implications for child language research: Comment on Tomasello. *Journal of Child Language* 25.462–66.

Bates, Elizabeth. 1999. Plasticity, localization and language development. *The changing nervous system: Neurobehavioral consequences of early brain disorders*, ed. by Sarah Broman and Jack M. Fletcher, 214–53. New York: Oxford University Press.

Blevins, James P. 2005. Word-based declensions in Estonian. *Yearbook of Morphology* 2005.1–25.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42.531–73.

Blevins, James P. 2007. Conjugation classes in Estonian. *Linguistica Uralica* 43.250–67.

Blevins, James P. 2013a. Word-based morphology from Aristotle to modern WP (word and paradigm models). *The Oxford handbook of the history of linguistics*, ed. by Keith Allan, 375–96. Oxford: Oxford University Press.

Blevins, James P. 2013b. The information-theoretic turn. *Psihologija* 46, to appear.

Blevins, James P. 2013c. *Word and paradigm morphology*. Oxford: Oxford University Press, to appear.

Blevins, James P., and Juliette Blevins. 2009. *Analogy in grammar: Form and acquisition*. Oxford: Oxford University Press.

Blumberg, Mark S. 2009. *Freaks of nature: What anomalies tell us about development and evolution*. Oxford: Oxford University Press.

Bochner, Harry. 1993. *Simplicity in generative grammar*. Berlin: Mouton de Gruyter.

Bonami, Olivier. 2013. Towards a robust assessment of implicative relations in inflectional systems. Paper presented at the Workshop on Computational Approaches to Morphological Complexity, Paris. Online: http://www.llf.cnrs.fr/Gens/Bonami/presentations/Bonami-SMG-Paris-2013.pdf.

Bonami, Olivier, and Gilles Boyé. 2007. French pronominal clitics and the design of paradigm function morphology. *Proceedings of the 5th Mediterranean Morphology Meeting*, 291–322. Online: http://mmm.lingue.unibo.it/mmm-proc/MMM5/MMM5-Proceedings_full.pdf.

Bonami, Olivier, and Fabiola Henri. 2010. Assessing empirically the inflectional complexity of Mauritian Creole. Paper presented at the Workshop on Formal Aspects of Creole Studies, Berlin. Online: http://www.llf.cnrs.fr/Gens/Bonami/presentations/BoHen-FACS-10.pdf.

Braine, Martin. 1966. Learning the positions of words relative to a marker element. *Journal of Experimental Psychology* 72.532–40.

Braine, Martin. 1987. What is learned in acquiring word classes: A step toward an acquisition theory. *Mechanisms of language acquisition*, ed. by Brian MacWhinney, 65–87. Hillsdale, NJ: Lawrence Erlbaum.

Bresnan, Joan. 1998. Morphology competes with syntax: Explaining typological variation in weak crossover effects. *Is the best good enough? Optimality and competition in syntax* (MIT working papers in linguistics), ed. by Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis, and David Pesetsky, 59–92. Cambridge, MA: MIT Press.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Philadelphia: John Benjamins.

Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82.711–33.

Capen, Carol Jamieson. 1996. *Diccionario Mazateco de Chiquihuitlán, Oaxaca*. (Serie de vocabularios y diccionarios indígenas 'Mariano Silva y Aceves' 34.) Tucson: Summer Institute of Linguistics.

Carstairs, Andrew. 1987. *Allomorphy in inflexion*. London: Croom Helm.

Carstairs-McCarthy, Andrew. 1991. Inflection classes: Two questions with one answer. *Paradigms: The economy of inflection*, ed. by Frans Plank, 213–53. Berlin: Mouton de Gruyter.

Carstairs-McCarthy, Andrew. 1994. Inflection classes, gender, and the principle of contrast. *Language* 70.737–88.

Carstairs-McCarthy, Andrew. 2010. *The evolution of morphology*. Oxford: Oxford University Press.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cinque, Guglielmo, and Luigi Rizzi. 2008. The cartography of syntactic structures. *CISCL Working Papers on Language and Cognition* 2.43–59. Online: http://www.ciscl.unisi.it/doc/doc_pub/cinque-rizzi2008-The_cartography_of_Syntactic_Structures.pdf.

Clark, Eve V. 1993. *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Davison, Anthony C., and David V. Hinkley. 1997. *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

De Vries, James A., and Sandra A. De Vries. 1997. An overview of Kwerba verb morphology. *Papers in Papuan Linguistics* 3.1–35.

Dobrin, Lise. 1999. *Phonological form, morphological class, and syntactic gender: The noun class systems of Papua New Guinea Arapeshan*. Chicago: University of Chicago dissertation.

Donohue, Mark. 2001. Animacy, class and gender in Burmeso. *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton* (Pacific linguistics 514), ed. by Andrew Pawley, Malcolm Ross, and Darrell Trevor Tryon, 97–115. Canberra: Australian National University.

Embick, David, and Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39.1–53.

Esper, Erwin A. 1973. *Analogy and association in linguistics and psychology*. Athens: University of Georgia Press.

Evans, Nicholas, and Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32.429–92.

Fortescue, Michael. 2002. The rise and fall of polysynthesis in the Eskimo-Aleut family. *Problems of polysynthesis*, ed. by Nicholas Evans and Hans-Jürgen Sasse, 257–76. Berlin: Akademie.

Frank, Wright J. 1999. Nuer noun morphology. Buffalo: State University of New York at Buffalo master's thesis.

Gerken, Lou Ann. 2009. The relation between linguistic analogies and lexical categories. In Blevins & Blevins, 101–17.

Gervain, Judit, and Ramon Guevara Erras. 2012. The statistical signature of morphosyntax: A study of Hungarian and Italian infant-directed speech. *Cognition* 125.263–87.

Gervain, Judit, and Janet F. Werker. 2013. Learning non-adjacent regularities at age 0;7. *Journal of Child Language* 40.4.860–72.

Griffiths, Paul. 2012. Our plastic nature. *Transformations of Lamarckism: From subtle fluids to molecular biology*, ed. by Snait B. Gissis and Eva Jablonka, 319–30. Cambridge, MA: MIT Press.

Hay, Jennifer, and R. Harald Baayen. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Science* 9.342–48.

Hein, Johannes, and Gereon Müller. 2009. Quantitative and qualitative aspects of paradigm economy in lesser studied languages. Paper presented at Morphology of the World's Languages, Universität Leipzig. Online: http://www.uni-leipzig.de/~muellerg/mu238.pdf.

Hockett, Charles F. 1987. *Refurbishing our foundations*. Amsterdam: John Benjamins.

Holman, Eugene. 1984. *Handbook of Finnish verbs*. Helsinki: Soumalaisen Kirjallisuuden Seuran.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51.639–71.

Jakobi, Angelika. 1990. *A Fur grammar: Phonology, morphophonology, and morphology*. Hamburg: Helmut Buske.

Jamieson, Carol. 1982. Conflated subsystems marking person and aspect in Chiquihuitlán Mazatec verbs. *Language* 48.139–67.

Julien, Marit. 2002. *Syntactic heads and word formation*. Oxford: Oxford University Press.

Juola, Patrick. 1998. Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics* 5.206–13.

Juola, Patrick. 2007. Assessing linguistic complexity. *Language complexity: Typology, contact, change*, ed. by Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson, 89–108. Amsterdam: John Benjamins.

Karmiloff-Smith, Annette. 1994. Transforming a partially structured brain into a creative mind. *Behavioral and Brain Sciences* 17.732–45.

Karmiloff-Smith, Annette. 1998. Development itself is the key to understanding developmental disorders. *Trends in Cognitive Science* 2.389–98.

Kibrik, Aleksandr E. 1998. Archi. *The handbook of morphology*, ed. by Andrew Spencer and Arnold M. Zwicky, 455–76. Oxford: Blackwell.

Kutsch Lojenga, Constance. 1994. *Ngiti: A Central-Sudanic language of Zaire*. Cologne: Rüdiger Köppe.

Lupyan, Gary, and Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS One* 5.e8559.

Malouf, Robert, and Farrell Ackerman. 2011. The low entropy conjecture: The challenges of Modern Irish nominal declensions. Paper presented at the LSA Workshop on Challenges of Complex Morphology to Morphological Theory, Boulder.

Matthews, Peter H. 1974. *Morphology*. Cambridge: Cambridge University Press.

Matthews, Peter H. 1991. *Morphology*. 2nd edn. Cambridge: Cambridge University Press.

Matthews, Peter H. 2001. *A short history of structural linguistics*. Cambridge: Cambridge University Press.

Milin, Petar; Victor Kuperman; Aleksandar Kostić; and R. Harald Baayen. 2009. Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. In Blevins & Blevins, 214–52.

Miller, George A. 1953. What is information measurement? *American Psychologist* 8.3–11.

Mintz, Toben H. 2002. Category induction from distributional cues in an artificial language. *Memory and Cognition* 30.678–86.

Moscoso del Prado Martín, Fermín. 2003. *Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies*. Nijmegen: Max Planck Institute for Psycholinguistics, Nijmegen dissertation.

Moscoso del Prado Martín, Fermín. 2011. The mirage of morphological complexity. *Proceedings of the 33rd annual conference of the Cognitive Science Society*, 3524–29.

Moscoso del Prado Martín, Fermín; Aleksandar Kostić; and R. Harald Baayen. 2004. Putting the bits together: An information theoretical perspective on morphological processing. *Cognition* 94.1–18.

Müller, Gereon. 2007. Morphological underspecification derives paradigm economy effects. Paper presented at the Workshop on Theoretical and Computational Perspectives on Underspecification, Stuttgart.

Oudeyer, Pierre-Yves. 2006. *Self-organization in the evolution of speech*. Oxford: Oxford University Press.

Paul, Hermann. 1891. *Principles of the history of language*. Trans. from 2nd edn. into English by Herbert A. Strong. London: Longmans, Green and Co.

Paunonen, Heikki. 1976. Allomorfien dynamiikkaa [The dynamics of allomorphs]. *Virittäjä* 79.82–107.

Ralli, Angela. 1994. Feature representations and feature-passing operations in Greek nominal inflection. *Proceedings of the 8th Symposium on English and Greek Linguistics*, 19–46. Thessaloniki: Aristotle University of Thessaloniki.

Ralli, Angela. 2002. The role of morphology in gender determination: Evidence from Modern Greek. *Linguistics* 40.519–51.

Ramscar, Michael, and Melody Dye. 2009. Error and expectation in language learning: An inquiry into the many curious incidences of 'mouses' in adult speech. *Proceedings of the 31st annual conference of the Cognitive Science Society*, 485–90.

Ramscar, Michael; Daniel Yarlett; Melody Dye; Katie Denny; and Kirsten Thorpe. 2010. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34.909–57.

Rescorla, Robert A. 1988a. Behavioral studies of Pavlovian conditioning. *Annual Review of Neuroscience* 11.329–52.

Rescorla, Robert A. 1988b. Pavlovian conditioning: It's not what you think it is. *American Psychologist* 43.151–60.

Roberts, John R. 1987. *Amele*. London: Croom Helm.

Robins, Robert H. 1959. In defense of WP. *Transactions of the Philological Society* 58. 116–44.

Sampson, Geoffrey B.; David Gil; and Peter Trudgill (eds.) 2010. *Language complexity as an evolving variable*. Oxford: Oxford University Press.

Shannon, Claude. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27.379–423, 623–56.

Sims, Andrea D. 2010. Probabilistic paradigmatics: Principal parts, predictability and (other) possible particular pieces of the puzzle. Paper presented at the 14th International Morphology Meeting, Budapest.

Sims, Andrea D. 2011. Information theory and paradigmatic morphology. Paper presented at the LSA Workshop on Information-Theoretic Approaches to Linguistics, Boulder. Online: http://www.ling.ohio-state.edu/LSAinfotheory/presentations/Sims.pdf.

Stoll, Sabine; Balthasar Bickel; Elena Lieven; Netra P. Paudyal; Goma Banjade; Toya N. Bhatta; Martin Gaenszle; Judith Pettigrew; Ichchha Purna Rai; Manoj Rai; and Novel Kishore Rai. 2012. Nouns and verbs in Chintang: Children's usage and surrounding adult speech. *Journal of Child Language* 39.284–321.

Stump, Gregory. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, Gregory. 2005. Some criticisms of Carstairs-McCarthy's conclusions. *Yearbook of Morphology* 2005.283–303.

Stump, Gregory. 2006. A theory of heteroclite inflectional paradigms. *Language* 82.279–322.

Stump, Gregory, and Rafael Finkel. 2007. Principal parts and morphological typology. *Morphology* 17.39–75.

Stump, Gregory, and Rafael Finkel. 2009. Principal parts and degrees of paradigmatic transparency. In Blevins & Blevins, 13–54.

Stump, Gregory, and Rafael Finkel. 2013. *Morphological typology: From word to paradigm*. Cambridge: Cambridge University Press.

Thelen, Esther, and Elizabeth A. Bates. 2003. Connectionism and dynamic systems: Are they really different? *Developmental Science* 6.378–91.

Thymé, Ann. 1993. *A connectionist approach to nominal inflection: Paradigm patterning and analogy in Finnish*. San Diego: University of California, San Diego dissertation.

Thymé, Ann; Farrell Ackerman; and Jeffrey L. Elman. 1994. Finnish nominal inflection: Paradigmatic patterns and token analogy. *The reality of linguistic rules*, ed. by Susan D. Lima, Roberta Corrigan, and Gregory K. Iverson, 445–66. Amsterdam: John Benjamins.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. London: Harvard University Press.

Trosterud, Trond. 2004. *Homonymy in the Uralic argument agreement systems*. Helsinki: Suomalais-Ugrilainen Seura.

Wohlgemuth, Jan, and Michael Cysouw (eds.) 2010. *Rethinking universals: How rarities affect linguistic theory*. Berlin: Mouton de Gruyter.

Wurzel, Wolfgang Ullrich. 1989. *Inflectional morphology and naturalness*. Dordrecht: Springer.

Ackerman
Department of Linguistics
University of California at San Diego
9500 Gilman Drive
La Jolla, CA 92093-0108
[fackerman@ucsd.edu]

Malouf
Department of Linguistics and Asian/Middle Eastern Languages
San Diego State University
5500 Campanile Drive
San Diego, CA 92181-7727
[rmalouf@mail.sdsu.edu]

[Received 2 March 2011;
revision invited 28 November 2011;
revision received 16 August 2012;
accepted 28 January 2013]