How inflection class systems work: On the informativity of implicative structure

Jeffery R. Parker  
*Brigham Young University, jeff_parker@byu.edu*

Andrea D. Sims  
*Ohio State University*

Follow this and additional works at: [https://scholarsarchive.byu.edu/facpub](https://scholarsarchive.byu.edu/facpub)

Part of the Linguistics Commons

**Original Publication Citation**

**BYU ScholarsArchive Citation**
[https://scholarsarchive.byu.edu/facpub/6258](https://scholarsarchive.byu.edu/facpub/6258)

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
How inflection class systems work: On the informativity of implicative structure

Andrea D. Sims
The Ohio State University

Jeff Parker
Brigham Young University

Abstract. The complexity of an inflection system can be defined as the average extent to which elements in the system inhibit motivated inferences about the realization of lexemes’ paradigm cells. Research shows that systems tend to exhibit relatively low complexity in this sense. However, relatively little work has explored how structural and distributional aspects of the inflectional system produce this outcome. In this paper we use the tools of information theory to do so. We explore a set of nine languages that have robust inflection class systems: Palantla Chinantec, French, Modern Greek, Icelandic, Kadiwéu, Nuer, Russian, Seri, and Vôro. The data show that the extent to which implicative paradigmatic structure does work to minimize the complexity of the system differs significantly. In fact, the nine languages fall into three graph types based on their implicative structure. Moreover, low type frequency classes disproportionally contribute to the complexity of inflectional systems, but we hypothesize that their freedom to detract in this way may depend on the extent to which implicative structure is systemically important. We thus propose that the amount of ‘work’ done by implicative relations in structuring inflection classes should be considered a typological parameter.

Keywords. inflection classes, morphology, typology, information theory, complexity, implicative structure, paradigmatic relations, type frequency, cross-linguistic comparison
1. Introduction

Languages differ in the extent to which they employ inflection class structure. Inflection classes are not necessary elements of language structure – and not all languages have them. Moreover, if the (teleological) goal of a language system were to create only as much structure as is needed to enable functional, efficient communication, probably no languages would have inflection classes. After all, they represent a classic example of morphomic structure – a layer of structure that mediates between form and meaning, without bearing meaning directly (Aronoff 1994). They thus introduce some amount of complexity into the grammatical system that is not directly connected to communicative need.

This has raised the question of whether there are universal limits on the complexity of inflection class systems. Best known are Paradigm Economy and subsequent work by Carstairs(-McCarthy) that have sought to define cross-linguistic limits on the number of possible inflection classes (Carstairs 1983, 1987; Carstairs-McCarthy 1994, 2010). The core insight of this work is that inflection classes must be sufficiently distinct to allow speakers to determine the class membership of a given lexeme. Carstairs-McCarthy thus defines the problem as one of inference. What properties of the inflectional system allow the speaker to infer a given lexeme’s full set of inflected forms (which is to say, its class membership)?

Recently, focus has shifted away from complexity defined in terms of the absolute number of inflection classes and towards a notion of complexity that is rooted in implicative paradigmatic structure. This is reflected in the way that Stump & Finkel (2013) extend the problem of inferring inflection class membership. They define the complexity of an inflection class system as ‘the extent to which the system inhibits motivated inferences about a lexeme’s full paradigm of realized cells FROM SUBSETS OF ITS CELLS’ (Stump & Finkel 2013: 55, emphasis)
ours). They focus on set-theoretic notions of principal part sets—i.e. a set of realized cells from which a lexeme’s full inflection class membership can be determined. The concept of a principal part set is, by its very nature, concerned with implicative paradigmatic structure, so Stump & Finkel inherently define the problem of inferring inflection class membership in terms of paradigmatic implication. Somewhat similarly, Ackerman & Malouf (2013) use information-theoretic tools to ask how much surprisal is associated with the inflected form realizing one paradigm cell, given the form associated with another cell (conditional entropy), and define the complexity of an inflection class system in terms of average conditional entropy.

At the same time, there is a particular sense in which neither is directly concerned with quantifying the work done by implicative structure. By defining complexity in terms of principal parts, Stump & Finkel have no way to assess the complexity of the system absent implicative structure, and thus no direct way to quantify how much implicative structure contributes towards minimizing the complexity of the system. And while Ackerman & Malouf’s metric reflects the complexity of inflectional systems, they do not much explore the work being done by implicative structure to produce this outcome. Ultimately, both works are primarily concerned with the end result—the extent to which inferences about a lexeme’s inflected forms are licensed.

Like these previous works, we are interested in the informativity of the structural and distributional properties of the system. However, we seek to answer what is in some sense a more basic question, namely, how much is implicative structure important for motivating inferences about the inflection class that a lexeme belongs to? Wurzel (1989: 114) was a major proponent of the importance of implicative paradigmatic structure to inflection classes:

"The inflectional paradigms are, as it were, kept together by implications. There
are no paradigms (except highly extreme cases of suppletion) that are not based on implications valid beyond the individual word, so that we are quite justified in saying that inflectional paradigms generally have an implicative structure, regardless of deviations in the individual cases. Words, whose paradigms have the same implicative structure form uniform inflectional (sub-)classes. In this sense, implications constitute inflectional classes.

In this paper we ask whether this is equally true for all languages that employ inflection classes. In addition to contributing to our understanding of morphological generalization and inflectional productivity, this is a typological question in its own right, related to the space for structural variation that languages inhabit. We use information-theoretic tools to develop an approach and offer an answer based on a set of nine languages.

The core questions of the present paper are: To what extent is implicative (paradigmatic) structure informative about the inflected forms of lexemes? In other words, how much ‘work’ does implicative structure do in maintaining inflection class systems? Is it similarly informative in all languages? What is the balance between implicative structure and other potential sources of information about inflected forms, in particular, inflection class type frequency? And finally, does the importance of implicative structure for inflection class structuring differ within inflection classes systems, from one class to another? These specific questions are informed by larger questions having to do with how structural elements interact in the context of the global properties of inflectional systems, and whether there are stable configurations into which inflection classes systems are organized.

The paper is structured as follows. §2 reviews some previous information-theoretic (and
other) work on inflection class complexity and the role of implicative structure in organizing inflection classes. In §3 we then introduce the nine languages that we explore in this paper. In §4 we define two information-theoretic metrics that we use to explore inflectional structuring: IMPLICATIVE WORK and TYPE FREQUENCY WORK. The next two sections are a data-driven exploration of the nine languages according to these metrics; §5 compares across the languages and §6 looks at interactions between implicative structure and type frequency within languages. Finally, in §7 we offer some conclusions.

2. Inflection class complexity and implicative structure

2.1. Problematizing implicative paradigmatic structure

Languages differ idiosyncratically in how morphosyntactic values are mapped to morphophonological form, leading some morphologists to emphasize the importance of investigating individual inflection class systems on their own terms (Aronoff 1994; Maiden 2005). While this does not equate to an expectation that inflection class systems should vary without constraint, inflection class structure has been ripe for investigation from a perspective that focuses at least as much on the internal organization of individual systems as on cross-linguistic comparison. In part for this reason, and in part because of the history of argumentation about the role of paradigmatic structure in inflectional theory, cross-linguistic diversity in the importance of implicative structure for inflection class organization has not been sufficiently problematized.

We consider there to be overwhelming evidence for inflectional rules that operate on a paradigmatic dimension (whether over fully inflected forms or at a more abstract level of representation), reflecting the line of argumentation that has developed from Matthews (1972), Anderson (1992) and Stump (2001), among other works. One of the benefits of a
paradigmatically-oriented theory is that it can capture ways in which inflection classes are self-reinforcing structures with independent organizing principles, structured around implicative relations holding between paradigm cells. To take a single example, Maiden (2005) argues that in Romance languages, verbal stem alternants exhibit patterns of change, such as analogical extension based on paradigmatic patterning alone, which indicate that their paradigmatic distributions are not synchronically accidental. Rather, they suggest a generalization at the abstract level of the distribution of stems, apart from the particular forms that realize the stem alternants (the so-called N-, L- and U-patterns in Romance verbs). Maiden thus concludes that implicative relations holding between cells of the paradigm, and specifically between stem morphs, create inflection classes as self-reinforcing structures. This is only one example of many arguments in favor of a paradigm-oriented model of inflection.

At the most basic level, the reasoning behind paradigm-based models is that if implicative structure is needed for an adequate description of the inflectional structure of some languages, then paradigmatically-oriented mechanisms must be part of the universal ‘toolbox’ from which languages can draw, and part of the architecture of inflectional systems. We do not challenge this basic reasoning, and in fact find there to be robust evidence for paradigmatic structure in (at least some) inflectional systems. However, paradigmatically-oriented rules being part of the toolbox does not necessarily mean that all languages employ implicative paradigmatic structure to a significant degree, even if they have inflection classes. The question of whether there is cross-linguistic diversity (or commonality) in the utilization of implicative paradigmatic structure is just as important a question as whether paradigmatically-oriented rules are in the toolbox in the first place.
2.2. Implicative structure as a typological parameter

Recent work has looked at how the structure of inflectional systems contributes to overcoming what Ackerman et al. (2009) call the Paradigm Cell Filling Problem, and has extended in the direction of questions related to inflectional complexity.² This includes studies of cross-linguistic differences and similarities in the distributions of inflectional systems (Ackerman & Malouf 2013; Stump & Finkel 2013), and in-depth studies of individual languages (e.g., Baerman (2012) for Nuer, Bonami (2013) for French, Sims (2015) for Greek, and Parker (2016) for Russian).

In (1) we give our definition of the complexity of an inflection class system. (This is similar to Stump & Finkel’s (2013) definition, cited in §1 above, but is more neutral with regard to the elements of the system that inhibit motivated inferences.)

(1) Complexity of an inflection class system: The average extent to which elements in an inflectional system inhibit motivated inferences about the realization of lexemes’ paradigm cells.

The complexity of an inflection class system can be operationalized using information-theory; a common metric is average conditional entropy (e.g., Ackerman & Malouf 2013). Conditional entropy H(A|B) represents the average surprisal associated with the outcome of a random variable A, given knowledge of the outcome of another random variable B. In the present context, A and B are paradigm cells in which A ≠ B. Implicitly conditioned on the lexeme, the outcomes of A and B are two inflected forms of the same lexeme. Conditional entropy thus represents the surprisal associated with an inflected form a realizing a given paradigm cell A, knowing another inflected form b of the same lexeme realizing paradigm cell B.
(2) Conditional Entropy

\[
H(A|B) = - \sum_{a \in A, b \in B} p(b, a) \log \frac{p(b)}{p(b, a)}
\]

The conditional entropy of A given B will never be higher than the entropy of A and will be lower whenever the value \(b\) is informative about the value \(a\). Knowing one form of a lexeme cannot increase the surprisal associated with another form, but it can lower it. Averaging across the weighted conditional entropy of all pairwise combinations of cells A and B, for all values of A and B in all inflection classes, produces an estimate of the complexity of the inflectional system as a whole.

In a survey of ten languages, Ackerman & Malouf (2013) find that the average surprisal associated with an inflected form given a knowledge of another form of the same lexeme (i.e. the average conditional entropy of the inflectional system) is uniformly relatively low, despite diversity in the size of the languages’ inflectional systems. They focus on the idea that implicative structure allows even large systems to exhibit low average conditional entropy and present their results as a typological tendency in the form of the Low Entropy Conjecture: ‘enumerative morphological complexity is effectively unrestricted, as long as the average conditional entropy, a measure of integrative complexity, is low’ (2013: 436). Stump & Finkel (2013: 215) offer a similar generalization in the form of the Depth-of-Inference Contrast: ‘languages show a high degree of uniformity in allowing a given form in a lexeme’s paradigm to be deduced from a low number of dynamic principal parts (the average number being not much more than one).’ Both approaches, thus, find evidence that even inflectional systems that vary widely in size exhibit uniformly low systemic complexity when it comes to the task of predicting
one inflected form from another.

Yet little direct attention has focused on whether there are typological generalizations about the ways in which systems attain low complexity. Stump & Finkel (2013) are centrally concerned with the role of implications in structuring inflectional paradigms. However, their set-theoretic principal parts approach cannot be used to define the predictability of an inflected form (or inflection class) absent implicative structure. This means that even in a hypothetical language in which there are no inflection classes and no competing allomorphs, the complexity of the inflectional system is still defined in terms of a principal part set. In such a case, any inflected form in the paradigm would by itself constitute a sufficient static or dynamic principal part set, but not because implicative structure is doing a lot of work in the system. In fact, its contribution is entirely vacuous. In the absence of competition between allomorphs, every inflected form of a lexeme is fully predictable, **INDEPENDENTLY** of any knowledge of the form realizing any other paradigm cell. Implicative structure does not improve the predictability of inflected forms. This fact is not directly captured by the set-theoretic principal parts approach. Ultimately, while Stump & Finkel’s metrics are suitable for describing the complexity of the system given the availability of implicative information, they do not directly measure the extent to which implicative structure is informative.

To substantiate the potential importance of implicative structure in attaining low average conditional entropy, Ackerman & Malouf run bootstrap simulations based on the inflection class systems of Chiquihuitlán Mazatec and Russian. These illustrate that implicative structure in Mazatec results in significantly lower average conditional entropy than would be expected from chance, but that in Russian it does not. They speculate that implicative organization is not important for maintaining low average conditional entropy in Russian. (Though see Parker
(2016) for a more detailed view of the inflectional properties of Russian nouns that comes to
different conclusions.) More generally, their argumentation suggests that implicative structure is
primarily important in large systems. But ultimately, Ackerman and Malouf are more focused
on the typological result of low average conditional entropy and pay little attention to the
different potential ways that systems may achieve this result.

Baerman (2012) represents a step down the path towards investigating the importance (or
lack thereof) of implicative structure in inflectional systems. In a study of inflectional exponence
in Nuer (a West Nilotic language spoken primarily in South Sudan and parts of Ethiopia),
Baerman argues that while the distribution of inflectional suffixes in the noun system exhibits
some degree of predictability, nonetheless the ‘paradigmatic distribution of Nuer case-number
suffixes is surprisingly unconstrained’ (474). Nuer nouns express three cases (nominative,
genitive, locative) and two numbers (singular, plural), but only a small set of formatives
competes to realize any given cell (one for nominative singular, two for plural, and three for
genitive singular and locative singular). Paradigm Economy and subsequent revisions of it
therefore predict a small number of inflection classes, yet Nuer has far more than this number, in
part because suffixes occur in nearly all of the possible combinations.

Baerman argues that implicative structure plays only a weak role in organizing inflection
classes of Nuer nouns; instead, relations between stems and suffixes within paradigm cells are
what matter. The distribution of stems is morphemic, and ‘there is no way to identify a stem as
singular or plural in isolation’ (Baerman 2012: 476). Moreover, the distribution of stems by case
is only mildly more constrained. But stems and endings exhibit weak complementarity: the
existence of stem alternation is correlated with zero suffixation, and non-alternation is correlated
with overt suffixation. The information about inflection class membership thus resides in the
predictive power of the distribution of stems. In the present context, the importance of Nuer lies
in the demonstration that low complexity need not derive from implicative structure holding
between paradigm cells, even in systems that are organized into inflection classes and which
have large paradigms or a large number of inflection classes.

We seek to expand on this work and fill in the gap by quantifying the amount of work
done by implicative paradigmatic structure in inflection class systems. Given that not all
languages have inflection classes to begin with, we think it is important not to presume that
paradigmatic relations necessarily plays much of a role. Even in inflection class systems, there
are other possible sources of information from which speakers may infer inflection class
membership. We thus see the ‘work’ done by implicative paradigmatic structure as a dimension
for comparison across languages, and a possible typological parameter.

2.3. Implicative structure within languages: The Marginal Detraction Hypothesis

We also ask whether inflection classes contribute equally to a system’s complexity. Stump &
Finkel (2013: 225) propose the Marginal Detraction Hypothesis, according to which ‘[m]arginal
I[nflection] C[lasse]s tend to detract most strongly from the IC predictability of other ICs’.
‘Marginal’ is defined in terms of the type frequency of the class – singleton and other low type
frequency inflection classes are marginal. Stump and Finkel show that Icelandic verbs are
consistent with the Marginal Detraction Hypothesis – singleton conjugations lower the inflection
class predictability of non-singleton conjugations more than the reverse (237). They treat this as
reflecting a historical tension between pressure towards maintaining singleton classes because
they are more predictable than non-singleton classes (presumably because they have high token
frequency, and are therefore less dependent on implicative paradigmatic structure), and the
pressure towards eliminating singleton classes because they detract from the predictability of
non-singleton classes. The hypothesis is thus that the relationship between regularity and frequency that has to do with the way that low type frequency classes participate in the morphomic (and especially, implicative) organization of the inflectional system. This intersects with work in discriminative models suggesting that there is a cognitive advantage in learning irregular phenomena (e.g., suppletion) because they are more easily discriminated, whereas there is an advantage towards generalizability for more regular patterns (Blevins et al. to appear a). This is indirectly supported by the fact that even in increasingly large corpora not all inflected forms are attested (Baayen 2001; Blevins et al. to appear b) and by evidence that the need to predict unknown forms remains relevant throughout the lifespan (Bonami & Beniamine 2015).

We seek to further explore the relationship between implicative structure and type frequency, as two possible sources of information for predicting inflected forms (although not the only possible ones), including testing a closely related idea to the Marginal Detraction Hypothesis.

3. Languages under investigation

In the remainder of the paper we investigate the properties of nine languages whose inflection class structure has recently been the topic of investigation: Palantla Chinantec, French, Modern Greek, Icelandic, Kadiwéu, Nuer, Russian, Seri, and Võro. We chose these languages because they offer a head-to-head comparison to previous work and because they belong to various language families. The sample is not meant to be representative of the world’s languages. Nonetheless, the data illustrate an interesting range of variation.

Palantla Chinantec is an Oto-Manguean language spoken in Oaxaca, Mexico (iso 639-3: cp). Verbs exhibit six tenses and four persons. Tone-stress patterns distinguish three stems in the paradigm; each stem realizes two tenses. Inflected forms are said to be predictable from the
tone-stress pattern, which is distributed over the three stems and four persons. Merrifield (1968) also notes that verbs exhibit paradigmatic patterns of stem segments and (ballistic) stress; we do not consider these. The data set we use is based on the 101 tone-stress patterns in the appendix of Merrifield & Anderson’s (2007) dictionary. A total of 838 word types were extracted from the digital version of the dictionary and manually checked.

French is a Romance (Indo-European) language spoken primarily in France (iso 639-3: fra). French verbs exhibit a variety of distinctions for person, number, mood and tense. Together these result in 49 paradigm cells in the synthetic paradigm. We use Stump & Finkel's (2013) hearer-oriented plat, with 72 classes of verbs based on 19 stems and 6485 total types, which is a (re)analysis of the verbal paradigms found in Bescherelle (2006). The complexity of French verbs is found in the stems (see Bonami & Boyé (2003) for discussion).

Modern Greek is a Hellenic (Indo-European) language spoken primarily in Greece (iso 639-3: ell). Greek nouns inflect for three cases and two numbers (plus a vocative of questionable status, which we leave out). The classes include differences in endings, inflectional stress, and stem changes. The data set we use here is based on the analysis in Sims (2015), which is derived from the Lexikó tís koinís neoellínikís (Triantafillidis Institute 1998). The data set contains 25,370 total types, falling into 49 classes.

Icelandic is a Northern-Germanic (Indo-European) language spoken in Iceland (iso 639-3: isl). Icelandic verbs exhibit distinctions for person, tense and mood, which result in 30 cells in the paradigm. We use the classes given by Stump & Finkel (2013) based on Jörg (1989), reflecting 162 classes with 1034 total word types.

Kadiwéu is a Mataco–Guaicuru language spoken in Brazil (iso 639-3: kbc). Kadiwéu verbs inflect for three persons and two numbers. There is no number distinction in the 2nd
person, resulting in five paradigm cells. Our data set is based on the verbal data from the Surrey Complexity Database (Baerman et al. 2015), based on Griffiths (2002). It has 61 classes containing 364 word types, with type frequencies based on the prefixes, stems and suffixes in the database.

Nuer is a Western Nilotic language spoken in South Sudan and parts of Ethiopia (iso 639-3: nus). Nuer nouns inflect for three cases and two numbers. The classes and type frequencies used here are taken from Table 7 in Baerman (2012: 470), reflecting 25 classes and 252 total word types in Frank (1999).

Russian is an East Slavic (Indo-European) language spoken in Russia and some neighboring countries (iso 639-3: rus). Russian nouns inflect for six main cases and two numbers. The data we use here are based on unique sets of distinguishers when including all suffixes, stem changes and stress patterns (see Parker 2016: Ch. 3 for details). Our data set represents 87 noun inflection classes and 43,486 total noun types, based on Zaliznjak (1977).

Seri is an isolate language spoken in Sonora, Mexico (iso 639-3: sei). Seri verbs realize two persons for subject and two aspects, resulting in a four-cell paradigm. Seri verbs also inflect for at least subject, object, tense, mood, negation and passive; however, the allomorphy among exponents for these properties is phonologically predictable. We use a set of 254 classes and 952 word types from the suffix data in Baerman (to appear), which is based on Moser & Marlett (2010). We only consider the distribution of suffixes in our set of classes.8

Võro is a Finnic language spoken in Estonia (iso 639-3: vro). It is either described as a variety of South Estonian or as an independent language. Võro verbs exhibit distinctions for at least person, number, tense, mood and voice. We adopt the analysis from Baerman (2014: 5, Table 3), which includes 23 classes, based on an analysis of data from Iva (2007). The data set
contains 4,668 total word types. The table he gives represents classes of verbs based on suffixes in eight cells from which the remainder of the paradigm can be predicted.\(^9\)

### 4. Implicative work and type frequency work

In this section we introduce the information-theoretic measures that we will use to explore inflection class structure. We define the notion of **implicative work** and explain how we operationalize it. Since we are interested in the interplay of different sources of information, we also define a notion of **type frequency work** and operationalize it.\(^{10}\)

#### 4.1. Inflection class complexity

An inflection class system’s complexity is a prerequisite notion to the idea of implicative work and type frequency work. We define the complexity of an inflection class system as in (1) above. We operationalize the inflection class complexity in terms of conditional entropy (see (2) above). To take account of the fact that within an inflectional system some inflection classes tend to have many more lexemes than others, we weight the component probabilities in our entropy calculations by the type frequencies of the possible outcomes of A and B (except where noted). The type frequency distribution of A (or B) is a function of the type frequencies of inflection classes and the extent to which its inflectional exponents occur in multiple classes. The weighted entropy of A (unconditioned or conditioned on B) will never be higher than the corresponding unweighted entropy of A and will be lower whenever the type frequency distribution of A is informative about the probability of \(a\). If all classes were equiprobable and all exponents occurred in the same number of classes, then type frequency would be entirely uninformative about inflection class membership. But in the typical case where inflection classes differ in frequency, and exponents may be shared across classes to different degrees, type frequency
allows for motivated inferences about forms – high type frequency exponents have higher a priori probability. We produce an estimate of the complexity of the inflectional system as a whole by averaging across the weighted conditional entropy values of all pairwise combinations of cells A and B.

4.2. Definitions of ‘work’

Average conditional entropy estimates the complexity of an inflectional system but it does not directly reflect the contribution of either implicative paradigmatic structure or type frequency to complexity. Inflectional systems can exhibit low conditional entropy in the absence of implicative structure. Reconsider the hypothetical language in which there is exactly one form that realizes each morphosyntactic property set, which is to say, there is no competition between allomorphs. In such a language, knowing a form b realizing paradigm cell B does not reduce the surprisal associated with the realization of paradigm cell A of the same lexeme because A is fully predictable to begin with (i.e., there is no surprisal associated with its outcome). This language would be consistent with the Low Entropy Conjecture, but implicative structure contributes vacuously to minimizing the complexity of the system. In order to investigate the role of implicative structure and type frequency directly, it is therefore necessary to establish a baseline against which the conditional entropy values can be compared.

The contribution of implicative paradigmatic structure, what we will call implicative work, thus has a natural formulation in terms of the extent to which knowledge of b as the outcome of B reduces the surprisal associated with the outcome of A.
Implicative Work

(a) The amount of reduction in the entropy of a paradigm cell A as a result of implicative relations holding among A and one or more other cells B in the same paradigm.

(b) $I(A;B)_U = H(A)_U - H(A|B)_U = H(B)_U - H(B|A)_U$

We operationalize implicative work as mutual information, also known as entropy reduction. The entropy $H(A)_U$ represents the entropy of A in the absence of any knowledge of B and without knowledge of the type frequency distribution of inflection classes – its unweighted, unconditioned entropy. The difference between $H(A)_U$ and $H(A|B)_U$ is the average extent to which the outcome of B is informative about the outcome of A (in the unweighted condition). Since mutual information is symmetric, the amount that knowing the value of A reduces the entropy of B is the same as the amount that knowing the value of B reduces the entropy of A. The mutual informativity of A and B, $I(A;B)_U$, thus captures the difference in the estimated complexity of the system before and after pairwise implicative paradigmatic relations are taken into account. This difference reflects the amount of work that is being done by implicative structure, in the absence of type frequency, to minimize the complexity of the inflectional system.

Likewise, we can measure the contribution of type frequency to minimizing the complexity of an inflection class system by calculating the extent to which the frequency distribution of inflection classes reduces the surprisal associated with a paradigm cell A.
(4) Type Frequency Work

(a) The amount of reduction in the entropy of a paradigm cell A as a result of the type frequency distribution of A.

(b) \( I(A) = H(A)_U - H(A)_W \)

We calculate type frequency work \( I(A) \) by calculating the (unconditioned) entropy without weighting probabilities by the type frequency of the classes, \( H(A)_U \), and subtracting the entropy that results when probabilities are weighted by frequency, \( H(A)_W \). The difference reflects the amount of work that is being done by type frequency, in the absence of implicative structure, to minimize the complexity of the inflectional system.

Of course, we can calculate the combined work done by both implicative structure and type frequency – the total amount of work – as the difference between \( H(A)_U \) and \( H(A|B)_W \). Moreover, implicative structure and type frequency may overlap in the work that they do, or otherwise interact within an inflectional system. We return to this issue below.

We calculate the average implicative work and/or average type frequency work being done in the inflectional system by averaging across all values of \( I(A;B) \) and \( I(A) \), respectively, for all paradigm cells and all inflection classes.

5. Cross-linguistic differences in implicative work and type frequency work

Turning to the structural properties of the nine languages described above, the first thing that we can observe is that the entropy of each language’s inflection class system is more similar when the work done by implicative structure and type frequency is included. Figure 1 shows that the difference among languages in weighted conditional entropy is much smaller than the difference
among languages in unweighted, unconditioned entropy. This is consistent with the Low Entropy Conjecture. More generally, the cross-linguistic similarities in weighted conditional entropy suggest that low complexity is a systemic property of inflectional systems. Like previous work, we hypothesize that it arises from speakers’ need to be able to make reliable inferences about inflected forms, the Paradigm Cell Filling Problem, mediated by reanalyses (analogical changes) that speakers make when the inflected forms of some lexeme are insufficiently predictable. We thus see low complexity as reflecting an evolutionary process; it emerges from an interaction between the individual elements in the system, in the environment of cognitive mechanisms for recall and productive generalization.

[Insert Figure 1 about here.]

Figure 1: Inflection class complexity

At the same time, Figure 1 also illustrates that the elements in the system that enable low complexity can differ substantially from one language to another – and this is what we are particularly interested in. Figure 2 extracts the amount of work done to minimize the complexity of each language’s inflection class system. (The bars in Figure 2 represent the distance between the two entropy values in Figure 1.) The values are given in Table 2. They show that implicative structure and type frequency combined do substantial work to minimize the complexity of the inflectional systems of French, Icelandic, and Greek. In Greek this is equivalent to reducing the choice among possible exponents for a paradigm cell by 11.23 (=$2^{3.49}$) forms on average. By way of contrast, in Võro and Nuer, the combined work of implicative structure and type frequency is quite modest. Entropy reduction in these languages corresponds to a reduction of 1.3 (=$2^{0.38}$) and 1.23 (=$2^{0.3}$) choices of exponent on average, respectively. As
Baerman (2012: 472) concludes for Nuer, this indicates that ‘the implicational network between the cells in the paradigm is weak’.  

Moreover, type frequency distributions and implicative structure do different amounts of work, both in absolute terms and in relative terms. For example, implicative structure reduces the entropy of the Greek nominal system by 3.27 bits (when type frequency information is not included), but only by 1.06 bits in the Russian nominal system. Conversely, type frequency distributions reduce the entropy of the Russian nominal system by 0.95 bits (when implicative structure is not included), but only 0.70 bits in Greek. In other words, type frequency does relatively more work in Russian, and implicative paradigmatic structure does relatively more
work in Greek. The data also show that across the group of nine languages, there are greater
differences in implicative work than in type frequency work.

5.1. Overlap in work

The data above show that implicative structure and type frequency do different amounts of work
in reducing uncertainty. Here we query the extent to which the work they do overlaps. For
example, type frequency does $x$ bits of work when implicative structure is ignored (difference
between unweighted and weighted unconditioned entropy) and $y$ bits of work when implicative
structure is included (difference between unweighted and weighted conditional entropy). The
difference between $x$ and $y$ is the overlap in work done by these information sources. Thus, the
overlap $W$ in work can be defined as a function of the difference between the work done by the
two information sources ($S_1$ and $S_2$) when considered alone, $I(S_1)$, and in the presence of the
other, $I(S_1|S_2)$.

\[
W(S_1;S_2) = I(S_1) - I(S_1|S_2)
\]

\[
= I(A;B)_U - I(A;B)_W
\]

This formula mirrors that of mutual information, i.e., $I(A;B) = H(A) - H(A|B)$, and shares the
property of being symmetric; $I(A;B) = I(B;A)$ and $W(S_1;S_2) = W(S_2;S_1)$. This measure offers
some insight into how type frequency and implicative structure interact in inflectional systems.
Overlap in work asks about the extent to which different sources of information about
inflectional exponence are redundant. From a functional perspective, we expect it to be
beneficial to speakers when inferences about inflectional realization are motivated in different
ways. This kind of redundancy may make the inflectional system robust to disruption and change based on imperfect learning. At the same time, as we discussed already in §2.3, there is reason to think that implicative structure and type frequency are not necessarily independent of each other.

[Insert Figure 3 about here.]

Figure 3: Overlap between implicative work and type frequency work (in bits)

Figure 3 illustrates that the overlap in type frequency and implicative work differs substantially across languages. Some languages exhibit significant overlap in the work done by each information source, but others exhibit little or even negative overlap. Negative overlap is particularly interesting. It occurs when implicative structure does more work when the underlying form probabilities are weighted by type frequency than it does when each inflectional exponent \( \{a_1, a_2, \ldots, a_n\} \) in the set of A is treated as equiprobable. It thus entails that implicative work is not evenly distributed across classes by type frequency. Specifically, weighting form probabilities by type frequency minimizes the influence of low type frequency classes. When those classes exhibit a disproportionately low level of inflection class predictability (i.e., they disproportionately contribute to the complexity of the system), this reduces positive work overlap and can produce negative overlap. We return to this issue in §6, where we talk about interactions between implicative structure and type frequency WITHIN inflectional systems.

5.2. Types of inflection class systems – a closer look at implicative distributions

The numbers reported in Table 2 and shown in Figures 2 and 3 are useful for getting a preliminary sense of differences in implicative and type frequency work, but by boiling down each language’s inflectional system to a single number, they necessarily obscure aspects of the
internal structuring of those systems. Here we borrow tools of graph theory to look in more detail at that structuring. A system with complete inflection class transparency (in the sense of Stump & Finkel 2013) has no exponents shared between classes – any form of any lexeme is sufficient to predict that lexeme’s inflection class membership, and would therefore by itself constitute an adequate principal part set. By contrast, a language with substantial overlap in exponents between classes exhibits some degree of inflection class complexity. The degree of complexity depends on the extent of the overlap and also factors like whether classes can be organized hierarchically or are cross-cutting (all else being equal, the latter results in greater complexity; see Sims 2015: Ch. 5 for some discussion).

[Insert Figure 4 about here.]
Figure 4: Network graphs of inflection class structure

Figure 4 visualizes inflection class structure as an undirected graph. The nodes are inflection classes, with the size of the node reflecting log type frequency of the class. An edge connecting two classes indicates that the classes share exponents. Black edges connect class nodes that share at least half of their exponents (e.g., for Russian, if at least 6 of 12 cells have the same inflectional exponent). Gray edges are classes that share half-minus-one cells (e.g., 5 cells in Russian). In order to facilitate visualization and keep the graphs from becoming too cluttered, we pruned edges between classes that share fewer exponents than this.

Visualized in this way, what emerges is the impressionistic observation that the nine languages represent three different kinds of inflection class systems. French, Greek and Icelandic are disconnected graphs, with relatively sparse overlap between classes – indicating strong
implicative structure. (Since strength of implicative structure is inversely related to amount of overlap between classes, amount of implicative work is inversely related to graph connectivity.) Palantla Chinantec, Kadiwéu and Russian, by comparison, are each (almost) connected graphs, and exhibit a high clustering coefficient. They resemble small world networks. Finally, Nuer, Seri and Võro exhibit strong connections between nodes and Nuer and Võro in particular are closer to being random graphs than the other languages are. The graphs thus show that implicative structure is not simply a scalar value – an inflectional system does not simply instantiate more or less of it. In graph-theoretic terms, the differences between the languages lie not only in connectivity, but also seemingly in the global clustering coefficient of each inflectional system network and other aspects of network structure (average topological distance, degree distribution). This suggests deeper differences in the structure of inflection class networks that are obscured by global calculations of inflectional complexity or implicative work.

5.3. Interim Summary

Returning to our major research questions, the data show clearly that the exponent of one paradigm cell can be highly informative about the exponents of other cells, but just as importantly, that the amount of work done by implicative structure (and to a lesser extent, by type frequency) varies enormously from one language to another, even within our small sample of languages. Thus, while the data are consistent with the idea that inflectional systems universally maintain low average conditional entropy, the distributional properties of individual elements that produce this result vary widely. Wurzel (1989: 114) said that ‘inflectional paradigms are, as it were, kept together by implications’, but this turns out not to be equally true for all inflection class systems, and substantially not true for some (Võro and Nuer in our data set). We therefore suggest the value of treating implicative paradigmatic structure as a
typological parameter of inflectional systems, describable in terms of the systems’ global network properties. Since this paper is primarily concerned with the relatively narrow issue of the relationship between inflectional complexity, implicative structure, and type frequency, we leave a full explication of graph-theoretic typological parameters for future work.

6. Implicative structure and type frequency within inflectional systems

Finally, we explore the relationship between implicative structure and type frequency within individual inflectional systems. Returning to an issue introduced in §2.3, the Marginal Detraction Hypothesis is an interesting hypothesis about the relationship between frequency and implicative structure, as sources of information about inflection class membership. This is the idea that low type frequency classes ‘tend to detract most strongly from the Inflection Class predictability of other ICs’ (Stump & Finkel 2013: 225). We explore here an idea that is a close cousin to the Marginal Detraction Hypothesis: Do some classes contribute more to the complexity of an inflectional system than others do? In particular, do low type frequency classes contribute more to the complexity of the system more than high type frequency classes do?

Our method involves calculating the difference between the average unweighted conditional entropy of the inflectional system as a whole and the average unweighted conditional entropy of the inflectional system with one class removed. We iterate over all inflection classes, and look at the effect of removing one class on the complexity of the system. If low type frequency classes disproportionately contribute to the complexity of the inflectional system, then we should find that the difference between the two entropies tends to be greater (and positive) when a low type frequency class is removed than when a high type frequency class is removed. In other words, there should be a negative correlation between inflection class type frequency and entropy difference. Entropy difference reflects the amount of complexity that is contributed
by the removed class.

\[(6) \quad \text{Entropy difference} = \text{mean}(H(A|B)_U) \text{ for full system} - \text{mean}(H(A|B)_U) \text{ for subset}\]

[Insert Figure 5 about here.]

Figure 5: Contribution of individual classes to the complexity of the inflectional system, by type frequency (left: Seri verbs; right: Icelandic verbs)

Figure 5 shows the results for Seri verbs (left) and Icelandic verbs (right). The Icelandic verbs are the same dataset that Stump & Finkel (2013) use to formulate the Marginal Detraction Hypothesis. The graph shows a significant negative correlation between the log type frequency of the removed inflection class and the difference in the average conditional entropy of the system. Thus, on average, low type frequency classes induce a greater degree of complexity in the system as a whole than higher type frequency classes do. The same pattern is found in Seri, although notice that the Seri classes are more tightly clustered around the regression line. The Icelandic verbs are more scattered and type frequency has less predictive power in Icelandic.

While our methods are different than Stump & Finkel’s, these results are consistent with the Marginal Detraction Hypothesis and expand on it by showing that the effect is not restricted to a sharp division between marginal and non-marginal classes. There is a continuum of marginality based on type frequency. The lower the type frequency of the class, the more it tends to increase the complexity of the system.\textsuperscript{17} Moreover, Table 3 shows that the majority of languages in our data set (six of nine) exhibit the same pattern. Only Greek shows a (non-significant) trend in the opposite direction. Thus, marginal detraction seems to be a tendency in
inflectional systems, though not a universal one.

<table>
<thead>
<tr>
<th>Language</th>
<th>Number of Classes</th>
<th>Total Types (lexemes)</th>
<th>Intercept</th>
<th>Slope</th>
<th>R²</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>72</td>
<td>6,485</td>
<td>0.0039</td>
<td>-0.00017</td>
<td>0.001</td>
<td>0.772</td>
</tr>
<tr>
<td>Icelandic</td>
<td>162</td>
<td>1,034</td>
<td>0.0023</td>
<td>-0.00097</td>
<td>0.067</td>
<td>0.001</td>
</tr>
<tr>
<td>Greek</td>
<td>49</td>
<td>25,370</td>
<td>0.0047</td>
<td>0.00097</td>
<td>0.038</td>
<td>0.180</td>
</tr>
<tr>
<td>Russian</td>
<td>87</td>
<td>43,486</td>
<td>0.0030</td>
<td>-0.00045</td>
<td>0.081</td>
<td>0.008</td>
</tr>
<tr>
<td>Kadiwêu</td>
<td>61</td>
<td>364</td>
<td>0.0069</td>
<td>-0.00200</td>
<td>0.127</td>
<td>0.005</td>
</tr>
<tr>
<td>Chinantec</td>
<td>101</td>
<td>838</td>
<td>0.0019</td>
<td>-0.00028</td>
<td>0.009</td>
<td>0.350</td>
</tr>
<tr>
<td>Seri</td>
<td>254</td>
<td>952</td>
<td>0.0017</td>
<td>-0.00088</td>
<td>0.272</td>
<td>0.000</td>
</tr>
<tr>
<td>Vôro</td>
<td>23</td>
<td>4,668</td>
<td>0.0118</td>
<td>-0.00179</td>
<td>0.140</td>
<td>0.079</td>
</tr>
<tr>
<td>Nuer</td>
<td>25</td>
<td>252</td>
<td>0.0017</td>
<td>-0.00088</td>
<td>0.301</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 2: Summary data for entropy difference as a function of type frequency

One interpretation is that low type frequency classes contribute disproportionately to the complexity of inflection class systems because their tendency to have high token frequency gives them a greater degree of independence from implicative structure. Having high token frequency makes the Paradigm Cell Filling Problem less of a problem, so there should be less need for implicative work. At the same time, it is not just the predictability of these classes that is at issue. They contribute to the complexity of their inflectional systems because they share one or more exponents with one or more other classes. So the predictability of the latter classes is also affected. This suggests that the pattern in Figure 5 is systemic in nature.

Exploring this issue further, we return to the discussion of negative overlap (Figure 3 in §5). The languages in our data set that exhibit very low or negative overlap in work – Vôro, Nuer, Kadiwêu and Chinantec – are also among those for which both overall work and implicative work are smallest (Figure 2 in §5). There is no inherent reason that this should be true, as far as we can see. However, it may indicate that in inflectional systems that overall make little use of implicative structure, low type (high token) frequency classes are free to exhibit a high degree of
independence. Specifically, they are free to exhibit a low degree of inflection class predictability and free to detract significantly from the predictability of forms in other classes. By comparison, in inflectional systems that rely heavily on implicative structure, even high token frequency classes may exhibit less independence. Even if implicative structure is not needed to solve the Paradigm Cell Filling Problem in those classes themselves, the systemic importance of implicative structure may limit the freedom of those classes to detract from the predictability of other classes.

Here we draw a parallel to morphological processing in English and Hebrew. Frost et al. (2000) show that in Hebrew masked visual priming, prime-target pairs sharing the same three-consonant root facilitate lexical retrieval of the target, even in the absence of semantic relatedness (e.g., klita ‘absorption’ – taklit ‘a record’, both of which are derived from the root klt yet share little if any semantic relationship). This contrasts with results for English (e.g., (Marslen-Wilson et al. 1994), where facilitation has been shown not to occur for semantically opaque prime-target pairs (e.g., successful – successor). Importantly, Hebrew words are highly structured according to a system of morphological templates. Plaut & Gonnerman (2000) found using connectionist modeling that in a morphologically rich artificial language (analogous to Hebrew), morphologically related but semantically opaque derived words were primed by their bases, whereas in a morphologically poor artificial language (analogous to English), no priming was found in this condition, although priming was found in more transparent items. They conclude that the morphologically rich language exhibited priming in the absence of semantic similarity because the organization of the internal representations in the network are dominated by the pervasive morphological structure of the language to such an extent that even opaque items participate in it. By contrast, in the impoverished language, the same items are free to
behave idiosyncratically’ (479). We speculate that the relationship between implicative structure and type frequency in inflectional systems may be similar in the sense that when implicative structure is a pervasive aspect of an inflectional system, even classes and lexemes for which it is not directly needed nonetheless are constrained by and participate in implicative structure, and are less free to behave idiosyncratically.

The fact that French and Greek do not exhibit marginal detraction is consistent with this interpretation, as highly implicative languages. Since the Icelandic verbal system exhibits a large amount of implicative work overall, we might not expect it to exhibit marginal detraction.

However, type frequency predicts only a small amount (6.7%) of the variance in entropy difference. (Compare Icelandic and Seri in Figure 5.) Moreover, the freedom to behave idiosyncratically that classes with low type frequency and high token frequency have is expected to be gradient. More research is thus needed, but we are inclined to interpret Icelandic as an intermediary point in gradient typological space.

7. Conclusions

Research on inflection class complexity exhibits a tension between two philosophical approaches: one that tends to emphasize language-particular idiosyncrasy and another that is interested in universal principles of organization. However, it is clear that we must investigate both simultaneously. Previous research shows that systems tend to exhibit relatively low complexity – on average, there is low uncertainty associated with unknown forms, making even large systems relatively simple for speakers. However, in this paper we showed that even in languages with robust inflection class systems, the extent to which implicative paradigmatic structure (and type frequency) do ‘work’ to reduce the complexity of the system differs significantly. As a result, implicative structure is much more important for structuring inflection
classes in some languages than in others. The idea that inflection is organized around paradigmatically-oriented implicative relations is central to the Word-and-Paradigm framework. As noted by Maiden (2005) and discussed above, inflection classes can become self-reinforcing structures, but we suggest that the implicative work done in such cases should be treated as being one point in the space of language variation. Cross-linguistic differences in the importance of implications for structuring inflection class systems has been insufficiently explored, but we have tried in this paper to give the problem some shape.

We also looked at how implicative paradigmatic structure interacts with other distributional elements within an inflectional system – in particular, inflection class type frequency. The most notable result here consists of evidence that in most of the languages in our sample, low type frequency classes contribute more to the complexity of the system than high type frequency classes do. This is consistent with the Marginal Detraction Hypothesis, although our metric was different. We suggest, albeit somewhat tentatively, that this may be a function of the extent to which the relevant inflectional system is reliant on implicative structure to license inferences about inflected forms of words and to maintain low complexity at a systemic level. In inflectional systems in which implicative structure plays less of a role, there may be more freedom for classes with high token frequency (which correlates with low type frequency) to behave idiosyncratically. In particular, such classes may be more free to detract from the predictability of forms in other classes. The distributional facts suggest the importance of investigating implicative structure not as an isolated element of inflectional systems, but as one that interacts with other sources of information in the context of the global properties of the system.

Ultimately, while most previous work has focused on the resultant complexity of the
system (whether in the form of Paradigm Economy, the Low Entropy Conjecture, etc.), the language-particular mechanisms by which low complexity is maintained offer an opportunity to better understand the organization of inflection class systems. We thus suggest that implicative work should be considered a typological parameter of inflection class systems.

References


http://www.smg.surrey.ac.uk/complexity/


Notes

1 We thank Matthew Baerman, Raphael Finkel and Greg Stump for providing data sets that we use in this paper. An early version of this work was presented at the International Quantitative Morphology Meeting (Belgrade 2015) and...
we thank the audience there for comments and ideas. This work was supported by the College of Arts and Sciences at Ohio State University through start up funds to the first author.


3 Stump & Finkel observe a difference in complexity between predicting one inflected form and predicting class membership (i.e., all forms). In contrast with the relatively uniform ease with which a single form can be deduced, ‘Languages vary widely in the number of dynamic principal parts they require to distinguish a given I[nflection] C[lass]’ (215). Ackerman & Malouf (2013: 443) likewise find greater cross-linguistic differences in average declensional entropy (an unconditioned entropy measure of inflection class predictability) than in average conditional entropy (a conditional entropy measure of inflected form predictability).

4 Sims (2011) also reasons that languages are likely to differ in the extent to which their inflectional systems are organized around implicative relations, based on the observation that in a language with small inflectional paradigms, each word-form will be (on average) encountered more often than in languages with large inflectional paradigms. Based on an investigation of Greek nouns, which have few paradigm cells, Sims (2015) shows that implicative organization can be important for small systems.

5 We are grateful to Matthew Baerman for sharing his data sets for Kadiwéu, Nuer, Seri and Vôro; and to Raphael Finkel and Greg Stump for sharing their data sets for French and Icelandic. The French data set was provided originally by Olivier Bonami, and we thank him as well. We built the Palantla Chinantec data set specifically for this paper, because data sets from previous work did not contain type frequency counts (Ackerman & Malouf 2013) and/or were drawn from older sources with a more limited representation of the system (Stump & Finkel 2013). The Modern Greek data set is described in more detail in Sims (2015). The Russian data set is described in more detail in Parker (2016).

6 Available online: http://morphologicaltypology.as.uky.edu/; file: complexity.french.data

7 Available online: http://morphologicaltypology.as.uky.edu/; file: complexity.icelandic.data

8 In addition to suffixes, some verbs exhibit paradigmatic patterns in the four cells based on a variety of changes in the stem, including variation in the stem final element, infixation, vowel syncope and ablaut.
As Baerman notes, the phonology of the stem and stem gradation can also restrict what suffix patterns are possible for a stem. We do not consider stem alternations or phonological restrictions on stem here.

There are also other kinds of work, including lexemic work. For instance, gender as a lexemic property is predictive of inflection class membership in some systems. Lexemic work (and its interaction with implicative work) is interesting. However, in this paper we leave this issue aside.

Unweighted unconditioned entropy is the same thing that Ackerman & Malouf (2013) call PARADIGM CELL ENTROPY.

We checked our script for calculating entropy values against the Principal Parts Analyzer (https://www.cs.uky.edu/~raphael/linguistics/analyze.html; accessed April 21, 2016). They produced the same values for all nine languages, including Nuer. We thank Raphael Finkel and Greg Stump for making this tool publicly available. Our script also produces the same unweighted conditional entropy values as reported in Ackerman & Malouf (2013) when run on their data structures for Russian and Greek. (These are the only languages that we tested in this way.) However, our script produces different numbers than Baerman (2012: 472) reports for Nuer. He lists an average unweighted unconditioned entropy value of 1.24 bits and an average weighted conditional entropy value of 0.87 bits for Nuer. (He calls these expected entropy and actual entropy, respectively.) As listed in Table 2, our script output 0.93 bits and 0.63 bits for the same calculations. Baerman reports that he used the Principal Parts Analyzer, and for our calculations we used the Nuer data set that he generously shared with us, so we cannot explain the discrepancy.

At the same time, we do not find evidence for Baerman’s claim (2012: 472) that weighted conditional entropy (ACTUAL ENTROPY in his terminology) is high in Nuer. In fact, Nuer has nearly the lowest weighted conditional entropy value among the languages in our data set. According to this metric, it is quite similar to the inflectional complexity of Icelandic and Greek. The crucial difference between the languages lies not in the complexity of their inflectional systems, but rather, in the amount of work done by implicative structure and type frequency distributions to minimize that complexity.

As is true of all entropy calculations, the output is highly sensitive to the nature of the input. For type frequency in particular, this means that the estimate of type frequency work is dependent on how representative the input type frequency distributions are. We feel confident that for Greek and Russian, we can make a good estimate of the type
frequency distribution of classes (and exponents in them), and correspondingly, the amount of work done by type frequency in each language. For the other languages we have made the best estimates of type frequency work and weighted conditional entropy that we can with the available data, but we acknowledge that having differing amounts of data makes it challenging to compare across languages.

15 The Kadiwéu data set contains 5 cells; we rounded down. So ‘half’ = 2 cells.

16 Unweighted conditional entropy is used to calculate entropy difference because the question addressed here is whether inflection class type frequency predicts entropy difference. So type frequency cannot be used also to calculate entropy difference.

17 We also find our particular formulation of the marginal detraction to be more intuitive, because it asks how individual classes contribute to the complexity of the system as a whole. Stump & Finkel’s version relies on a partitioning of the data into singleton and non-singleton classes, and a calculation of how each partition affects the predictability of the other.