3

Parts and wholes: Implicative patterns in inflectional paradigms

Farrell Ackerman, James P. Blevins, and Robert Malouf

The whole has value only through its parts, and the parts have value only by virtue of their place in the whole. (Saussure 1916: 128)

. . . we cannot but conclude that linguistic form may and should be studied as types of patterning, apart from the associated functions. (Sapir 1921: 60)

3.1 Introduction

This chapter addresses an issue in morphological theory – and, ultimately, morphological learning – that we feel has received far less attention than it deserves. We will refer to this issue as the **Paradigm Cell Filling Problem** (PCFP):

Paradigm Cell Filling Problem: What licenses reliable inferences about the inflected (and derived) surface forms of a lexical item?

The problem does not arise in an isolating language, in which each lexical item (or "lexeme") is realized by a single form. English, for all intents and purposes, approaches an isolating ideal, so that the PCFP has not been prominent in analyses of English (or, for that matter, in the post-Bloomfieldian morphological models that have been developed mainly within the English-speaking world).[1]

However, the PCFP arises in an acute form in languages with complex inflectional systems, especially those which contain large inflectional paradigms

[1] Though a concern with form and structure of paradigms has remained a central focus of other morphological traditions, as represented by Seiler (1965), Wurzel (1970), and Carstairs (1983).

and intricate inflection-class systems. For example, a typical Estonian noun paradigm contains 30-odd forms, which exhibit patterns of variation that place the noun within anywhere between a half-dozen and a dozen major declension classes (Viks 1992; Erelt *et al.* 1995; Blevins 2005). 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, Estonian is far from an extreme case. A typical transitive verb in Georgian has upwards of 200 forms, whose inflectional patterns identify the verb as belonging to one of four major conjugation classes (Tschenkéli 1958). Even Georgian is relatively conservative in comparison with descriptions of verb paradigms in Archi, which, according to one estimate (Kibrik 1998: 467), may contain "more than one and a half million" members.

The basic challenge that a speaker faces in each of these cases is the same, irrespective of the size of the form inventory. Given prior exposure to at most a subset of forms, how does a speaker produce or interpret a novel form of an item? One superficially attractive intuition is that knowing **what** one wants to say suffices in general to determine **how** one says it. The idea that variation in form reflects differences in "grammatical meaning" is encapsulated in the post-Bloomfieldian "morpheme," and underlies morphemic models from Harris (1942) and Hockett (1947) through Lieber (1992) and Halle and Marantz (1993). Yet, if one thing has been established about morphological systems in the half-century since Hockett (1954), it is that complex systems exhibit genuinely morphological variation, which is not conditioned by differences in grammatical meaning (or, for that matter, solely by phonological factors). Purely morphological variation (or what Aronoff 1994 terms "morphology by itself") may seem enigmatic in the context of simple systems. But in larger and more complex systems, variation that identifies the class of an item contributes information of vital importance because it allows a speaker to predict other forms of the item.

In a language with inflection classes, a speaker must be able to identify the class of an item in order to solve the PCFP. That is, to produce or interpret a novel form of an item, it is not enough for the speaker to know just that the grammatical meaning "motion into" is expressed by the illative case. The speaker must also know how the illative is realized for the item in question. In an inflection-class language, the choice of stem choice or exponent is precisely what is not in general determinable from the semantic or grammatical properties of an item. Instead, a speaker must know, or be able to deduce, one of the diagnostic forms of an item. For example, no known grammatical properties explain why the Estonian noun LUKK 'lock' has the short illative singular form *lukku* alongside the long form *lukusse*, whereas KIRIK 'church' has just the long form *kirikusse*. However, this contrast follows immediately if one knows that the partitive singular of LUKK is *lukku* and that the partitive singular of KIRIK is *kirikut*. There is likewise no morphosyntactic motivation for the variation in the form of the illative singulars in the Saami paradigms in Table 3.3 below. It is a morphological fact that the illative singular *bihttái* is based on the strong stem of BIHTTÁ 'piece' whereas the illative singular *bastii* is based on the weak stem of BASTE 'spoon.' This contrast is again predictable from the grade of the nominative singular forms of each noun (or, indeed, from the grade of any other form, as shown in Section 3.1).

In short, morphological systems exhibit interdependencies of precisely the kind that facilitate the deduction of new forms, based on knowledge of other forms. In some cases, it may be possible to mediate these deductions through a level of analysis in which recurrent units of form are associated with discrete grammatical meanings. However, this type of analysis tends to be most applicable to simple or recently grammaticalized patterns, and most morphological systems are not organized in a way that facilitates the identification of "minimal meaningful units". In many cases, the interdependencies that hold between word forms do not hold between subword units, so that further analysis disrupts the implicational structure. For example, the partitive singular *lukku* implies the homophonous short illative singular *lukku*, even though neither *lukk* nor -*u* can be associated with the grammatical meaning "partitive" or "illative" (Blevins 2005).

To develop this perspective, Section 3.2 outlines the word and paradigm assumptions that underlie our analysis, together with the basic information theoretic measures we use to test these assumptions. In Section 3.3, we apply these measures to portions of the morphological systems of Saami and Finnish and argue that – even in the absence of accurate frequency information – these measures bring out an implicational structure that offers a solution to the PCFP. We then show how the same measures apply to a description of Tundra Nenets nouns that supplies information about type frequency.[2] Taken together, these case studies suggest how information theory can be used to measure the implicational relations that underlie **symmetrical** approaches to word relatedness. By measuring the information that multiple surface forms provide about other forms, these approaches capture patterns of interdependency that cannot always be expressed in terms of an **asymmetrical** relation between surface forms and a single underlying or surface base.[3] We illustrate a symmetrical approach by examining Tundra Nenets nominal

[2] The fieldwork on Tundra Nenets was supported by a Hans Rausing Endangered Language Major Documentation Project Grant 2003–6, in which the first author was a co-PI with Irina Nikolaeva and Tapani Salminen. This support is gratefully acknowledged.

[3] See Albright (2002*a*, this volume) for a single base approach that addresses language change.

declension classes for absolute paradigms, and offer some provisional results about paradigm organization in this language. Section 3.4 then closes with some general conclusions and speculates about their ramifications for theoretical approaches to morphological analysis.

3.2 Analytical assumptions

Processes of analogical pattern matching and pattern extension play a central role in traditional analyses of interdependencies within and across paradigms. In classical word and paradigm (WP) models, a morphological system is factored into two components: a set of exemplary paradigms that exhibit the inflectional patterns of a language, and sets of diagnostic principal parts for nonexemplary items. Matching diagnostic forms of an item against the corresponding cells in an exemplary paradigm provides an analogical base for the deduction of novel forms of the item. This process of matching and deduction tends to be expressed symbolically in terms of proportional analogies (discussed in more detail in Albright (this volume) and Milin *et al.* (this volume)). The same process is invoked in grammars of inflectionally complex languages, as illustrated by the "rules of analogy" in Viks (1992: 46), which identify those forms of an Estonian noun that are predictable from the genitive singular and from the genitive plural.

3.2.1 *Morphological assumptions*

Traditional WP models offer a general solution to the PCFP that exploits the implicational structure of inflectional systems. Strategies that use exemplary patterns to extend principal part inventories are strikingly effective, as Matthews (1991: 187) notes in connection with their pedagogical relevance. They are also remarkably economical. In general, a small set of principal parts is sufficient to identify the class of an item and predict other forms of the item. Yet traditional solutions to the PCFP also raise some basic questions, including those in (1):

(1) a. What is the structure of units that license implicative relations?
b. How are units organized into larger structures within a system?
c. How can one measure implicative relations between these units?
d. How might the implicative organization of a system contribute to licensing inferences that solve the paradigm cell filling problem?
e. How does this organization, and the surface inferences it licenses, contribute to the robustness and learnability of complex systems?

Questions (1a) and (1b) centrally concern the relation of parts to wholes along two independent dimensions of analysis. Question (1a) concerns the internal complexity of word forms. Within post-Bloomfieldian models, words are treated as aggregates of smaller meaningful elements. These parts combine to produce a whole whose meaning is just the sum of the meaning its parts. Within the WP approach adopted here, words are regarded as complex configurations of recurrent elements whose specific **patterns of combination** may be meaningful irrespective of whether any particular piece bears a discrete meaning.

From this perspective, a surface word form is a whole in which the patterns exhibited by parts – whether affixes, tones, ablaut, or other "features of arrangement" (Bloomfield 1933: 163) – merely signal morphosyntactic, lexical, or morphological properties.[4] For example, in Tundra Nenets, the same members of a suffix set can be used with different lexical categories, sometimes serving essentially the same function, and sometimes serving different functions.[5]

As shown in Table 3.1, markers from Suffix Set I can appear both on nouns and verbs, and the inflected word functions as the predicate of the clause. In either case, the set I markers reflect person and number properties of the clausal subject. While markers from Suffix Set II also occur either with with nouns or with verbs, their function differs within each class: they reflect person/number properties of the possessor when they appear with nouns, but number properties of clausal objects when they appear with (transitive) verbs. Hence, there is a configurational dynamic whereby the same elements in different combinations are associated with different meanings. These patterns show why words are best construed as **recombinant gestalts**, rather

[4] This perspective does not preclude the possibility of associating grammatical meaning with subword units (morphemes) in constructions and/or languages where they would be motivated. In contrast, a morphemic model is less flexible, as it uniformly associates grammatical meaning with minimal elements and ignores configurational (emergent) properties of patterns.

[5] This discussion follows the presentation in Salminen (1997: 96, 103, 126), though here as elsewhere we have simplified his transcriptions for a general audience. In particular, we have largely rendered the traditional Cyrillic written conventions into an IPA-based system where digraphs such as *ty* indicate palatalized consonants, `refers to a glottal stop with nasalizing or voicing effects in sandhi contexts, and ``refers to a glottal stop without nasalizing effects in sandhi contexts. (For a detailed discussion of the phonological properties of words and how orthographic conventions reflect particular analyses, see Salminen (1993).) Also, while predicate nominals and adjectives in Tundra Nenets host markers from Suffix Set I, they differ from the verbal predicates that host these suffixes in exhibiting nominal stem formation rather than verbal stem formation, in the inability to host future markers, and in their manner of clausal negation. All of these differences suggest that two different lexical categories host markers from Suffix Set I, and that there is no N-to-V conversion operation.

Table 3.1 Suffix homonymy in Tundra Nenets

	N	V
Suffix Set I	Predicative	Subjective
Suffix Set II	Possessive	Objective

than simple (or even complex) combinations of bi-unique content-form mappings (i.e., morphemes).[6]

This perspective on complex words is intimated in Saussure (1916: 128) in his discussion of **associative** (= paradigmatic) relations:

A unit like **painful** decomposes into two subunits (**pain-ful**), both these units are not two independent parts that are simply lumped together (**pain** + **ful**), The unit is a product, a combination of two interdependent elements that acquire value only through their reciprocal action in a higher unit (**pain** × **ful**). The suffix is non-existent when considered independently; what gives it a place in the language is series of common terms like **delight-ful, fright-ful**, etc. . . . The whole has value only through its parts, and the parts have value by virtue of their place in the whole.

Accordingly, while we are often able to isolate pieces of complex form, it is the configurations in which these pieces occur and the relation of these configuration to other similar configurations that are the loci of the meanings that are relevant in morphology. This property becomes even more evident if one considers the structure of Tundra Nenets verbs as schematized in Salminen (1997).

Table 3.2 exhibits little in the way of a one-to-one correspondence between cells across columns. Consider first the general finite stem, whose use is exemplified in (2). This stem serves as the base for the subjective conjugation,

Table 3.2 More suffix homonomy in Tundra Nenets

Conjugation	Number of Object	Morphological Substem	Suffix Set
subjective		general finite stem (modal substem)	I
	sg		II
objective	du	dual object (modal) substem	III
	pl	special finite stem	
reflexive		special modal stem	IV

[6] See Gurevich (2006) for an constructional analysis of Georgian along these lines.

as shown in (2a), the objective conjugation, as shown in (2b), and may also encode singular object agreement for verbs marked by Suffix Set II. The dual object (modal) substem hosts members of Suffix Set III, as exemplified in (3), but the same suffix set also serves to mark plural objects with the special finite stem in (4a). Finally, as (4b) shows, the special finite stem is not restricted to the plural object conjugation, given that it is also associated with the reflexive conjugation and the distinguishing characteristic of this conjugation is the use of suffix set IV.

(2) General finite stem:
  a. Subjective:
    tonta-dm′
    cover.I (= 1sg)
    'I cover (something)'
  b. Objective Singular:
    tonta-w
    cover.II (= 1sg/sg)
    'I cover it'

(3) Dual Object Stem:
  tonta-gaxayu-n
  cover.dual.III (= 1sg/du)
  'I cover them (two)'

(4) Special finite stem:
  a. Objective Plural
    tontey-n
    cover.III (= 1sg/pl)
    'I cover them (plural)'
  b. Reflexive
    tontey-w″
    cover.IV (= 1sg)
    'I got covered'

In sum, it is the pattern of arrangements of individual elements that realize the relevant lexical and morphosyntactic content associated with words that is important in these examples, rather than the sum of uniquely meaningful pieces.

A word-based perspective on these aspects of the **internal** organization of lexical units is highly compatible with a traditional conception of the second part-whole dimension, namely the **external** organization of words. In what Matthews (1991) below terms the "ancient model", individual words function as minimal elements in networks of elements, including inflectional paradigms, and paradigms are organized into larger networks, which include inflection classes.

In the ancient model the primary insight is not that words can be split into formatives, but they can be located in paradigms. They are not wholes composes of simple parts, but are themselves the parts within a complex whole. (Matthews 1991: 204)

The notions of internal and external structure are not exclusive – as they are sometimes thought to be – but, instead, represent complementary perspectives on a morphological system. Indeed, these two dimensions give rise to a paradigmatic variant of "duality of patterning" (Hockett 1960), in that they show how combinations of individually meaningless elements, whether morphs or other "features of arrangement", compose words whose meaning depends in part on the place they occupy within larger paradigmatic structures. These complementary notions also permit an exploration of the intuitions evident in the twin themes of the epigrams above. In order to address these issues, the following sections explore how several Uralic languages (Saami, Finnish, and particularly Tundra Nenets) provide fertile ground for identifying the nature of the challenges posed by the PCFP, as well as the type of analysis best suited to address them.

Traditional WP approaches suggest answers to the other questions in (1), though in addressing these questions, it is important to separate the substantive claims and hypotheses of a WP model from any idealizations or simplifying assumptions introduced in the use of these models in reference or pedagogical grammars. For practical purposes, it is usually convenient in written grammars to represent lexical items by a single principal part wherever possible. Yet there is no reason to attribute any linguistic or psychological relevance to this extreme level of lexical economy. There are many well-described systems in which class can only be identified on the basis of multiple principle parts. Estonian conjugations provide a fairly straightforward illustration (Blevins 2007) as do the systems described in Finkel and Stump (this volume).[7] From a psycholinguistic perspective, there is considerable evidence that frequency is, in fact, the primary determinant of whether a given form is stored in the mental lexicon of a speaker (Stemberger and MacWhinney 1986; Baayen *et al.* 2003*b*). Similarly, grammars tend to take the smallest diagnostic forms of an item as principal parts, even though any form

[7] It may be significant that models incorporating something like the "single base hypothesis" (Albright 2002*a*, this volume) tend to be developed on the basis of comparatively simple systems.

(or set of forms) that identifies class is equally useful, and the choice of memorized forms is again likely to reflect frequency or other distributional properties rather than morphosyntactic or morphotactic properties.

Other issues that are implicit in traditional analogical models have been addressed in recent work. Methods for identifying and classifying principal part inventories are set out in Finkel and Stump (2007, this volume). The psychological status of proportional analogies is likewise addressed in Milin *et al.* (this volume). But traditional solutions to the PCFP remain fundamentally incomplete to the extent that they lack a means of gauging the diagnostic value of principal parts or of measuring the implicational structure of networks of forms.

The approach outlined in this paper proceeds from the observation that implicational structure involves a type of **information**, specifically information that forms within a set convey about other forms in that set. Information in this sense corresponds to reduction in **uncertainty**. The more informative a given form is about a set of forms, the less uncertainty there is about the other forms in the set. The PCFP just reflects the fact that a speaker who has not encountered all of the forms of a given item is faced with some amount of uncertainty in determining the unencountered forms. If the choice of each form were completely independent, the PCFP would reduce to the problem of learning the lexicon of an isolating language. However, in nearly all inflectional systems, there are at least some forms of an item that reduce uncertainty about the other forms of the item. It is the reduction in uncertainty due to the knowledge of these forms that defines the implicational structure of the system. The diagnostic value of a given form likewise correlates with the reduction in uncertainty that is attributable to the knowledge of this particular form. Once these notions are construed in terms of uncertainty reduction, the task of measuring implicational structure and diagnostic value is susceptible to well-established techniques of analysis.

3.2.2 *Information theoretic assumptions*

The uncertainty associated with the realization of a paradigm cell correlates with its **entropy** (Shannon 1948) and the entropy of a paradigm is the sum of the entropies of its cells. The implicational relation between a paradigm cell and a set of cells is modeled by **conditional entropy**, the amount of uncertainty about the realization of the set that remains once the realization of the cell is known. Finally, the diagnostic value of a paradigm cell correlates with the **expected conditional entropy** of the cell, the average uncertainty that remains in the other cells once the realization of the cell is known.

A straightforward application of these information-theoretic notions provides a natural means of measuring the implicational structure of inflectional systems. In particular, we use the notion of **information entropy** to quantify the uncertainty in the realization of a particular cell of a paradigm. As in Moscoso del Prado Martín *et al.* (2004), Milin *et al.* (2009) and Milin *et al.* (this volume), an information-theoretic perspective permits us to reconsider basic linguistic questions, in this case questions about the synchronic structure of inflectional systems.

In order to quantify the interrelations between forms in a paradigm, we use the information theoretic notion **entropy** as the measure of predictability. This permits us to quantify "prediction" as a change in uncertainty, or information entropy (Shannon 1948). The idea behind information entropy is deceptively simple: Suppose we are given a random variable X which can take on one of a set of alternative values $x_1, x_2, \ldots, x_n$ with probability $P(x_1), P(x_2), \ldots, P(x_n)$. Then, the amount of uncertainty in X, or, alternatively, the degree of surprise we experience on learning the true value of X, is given by the entropy $H(X)$:

$$H(X) = -\sum_{x \in X} P(X) \log_2 P(X)$$

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 1 bit is the information in a choice between two equally probable outcomes. Outcomes which are less probable (and therefore less predictable) have higher surprisal. Surprisal is 0 bits for outcomes which always occur ($P(x) = 1$) and approaches ∞ for very unlikely events (as $P(x)$ approaches 0). 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 will be made among competitors 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.

For example, the entropy of a coin flip as resulting in either heads or tails is 1 bit; there is equal probability for an outcome of either heads or tails:

$$\begin{aligned} H(X) &= -\sum_{x \in X} P(x) \log_2 P(x) \\ &= -(P(\mathrm{h}) \times \log_2 P(\mathrm{h}) + P(\mathrm{t}) \times \log_2 P(\mathrm{t})) \\ &= -(0.5 \times \log_2 0.5 + 0.5 \times \log_2 0.5) \\ &= 1 \end{aligned}$$

The entropy of a coin rigged to always come up heads, on the other hand, is 0 bits: there is no uncertainty in the outcome:

$$\begin{aligned} H(X) &= -\sum_{x \in X} P(x) \log_2 P(x) \\ &= -(P(\text{h}) \times \log_2 P(\text{h}) + P(\text{t}) \times \log_2 P(\text{t})) \\ &= -(1.0 \times \log_2 1.0 + 0.0 \times \log_2 0.0) \\ &= 0 \end{aligned}$$

For other possible unfair coins, the entropy will fall somewhere between these extremes, with more biased coins having a lower entropy. We can extend this to find the **joint entropy** of more than one random variable. In general, the joint entropy of independent events is the sum of the entropies of the individual events. Suppose *X* is the outcome of one flip of a fair coin and *Y* is the outcome of a second flip. If the two flips are independent, then the probability of getting, say, heads on the first flip and tails on the second is the probability of getting heads on first times the probability of getting tails on the second, or $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. So, then, the joint entropy $H(X, Y)$ is:

$$\begin{aligned} H(X,Y) &= -\sum_{x \in X, y \in Y} P(x,y) \log_2 P(x,y) \\ &= -(P(\text{h},\text{h}) \times \log_2 P(\text{h},\text{h}) + P(\text{h},\text{t}) \times \log_2 P(\text{h},\text{t}) \\ &\quad + P(\text{t},\text{h}) \times \log_2 P(\text{t},\text{h}) + P(\text{t},\text{t}) \times \log_2 P(\text{t},\text{t})) \\ &= -4 \times (0.25 \times \log_2 0.25) \\ &= 2 \end{aligned}$$

3.3 Modeling implicational structure

With the previous section as background we can now measure the entropy of the inflectional systems mentioned earlier. In order to exhibit the general character of the PCFP and demonstrate how an information-theoretic approach calculates the relative diagnosticity of words, the following subsections present several morphological patterns with ascending levels of complexity. We first describe the basic patterns, restricting attention to instructive aspects of the organization of these systems, and then develop entropy-based analyses that reveal their implicational structure. The inflectional paradigms of Uralic languages are particularly instructive because of the way that they realize inflectional properties by distinctive combinations of stem alternations and affixal exponence. Hence these systems are not amenable to a standard head-thorax-abdomen analysis in which lexical properties are expressed by the root, morphological class properties by stem formatives, and inflectional properties by inflectional affixes. For expositional convenience, we will initially assume, contrary to fact, that each cell in the paradigms below are equiprobable, so that speakers are just as likely to encounter one specific cell as any other.[8] As will be shown in the following sections, an appealing property of an entropy-based measure of word relatedness is that they can be easily scaled up to data sets of increasing veridicality.

3.3.1 *Northern Saami*

Noun declensions in Northern Saami (Bartens 1989; Nickel 1990) offer a straightforward illustration of the PCFP. First-declension nouns, i.e., nouns whose stems have an even number of syllables, may inflect according to either of the patterns in Table 3.3. In nouns of the "weakening" type, the nominative and illative singular and the essive are all based on the strong stem of a noun, and the remaining forms are based on the weak stem. Nouns of the "strengthening" variety exhibit a mirror-image pattern, in which the nominative and illative singular and essive are based on the weak stem, and other forms are based on the strong stem. Strong forms, which are set in bold in Table 3.3, contain a geminate consonant which corresponds to a nongeminate in the corresponding weak forms.

On standard descriptions that recognize a single, number-neutral essive form, there are eleven cells in a first-declension paradigm. Hence, to solve the PCFP, a speaker must deduce at most ten forms. This task is greatly facilitated

Table 3.3 Gradation in first declension nouns in Saami (Bartens 1989: 511)

	'Weakening'		'Strengthening'	
	Sing	Plu	Sing	Plu
Nominative	**bihttá**	bihtát	baste	**basttet**
Gen/Acc	bihtá	bihtáid	**bastte**	**basttiid**
Illative	**bihttái**	bihtáide	bastii	**basttiide**
Locative	bihtás	bihtáin	**basttes**	**basttiin**
Comitative	bihtáin	bihtáiguin	**basttiin**	**basttiiguin**
Essive		**bihttán**		basten
		'piece'		'spoon'

[8] Assuming equiprobable realizations also gives us an upper bound on the uncertainty in a paradigm. Since it is unlikely that all realizations are in fact equally likely, the actual entropy will almost always be lower than this.

Table 3.4 Invariant case endings in Saami (*e* assimilates to *i* before *i*)

	Sing	Plu
Nominative	—	-t
Gen/Acc	—	-id
Illative	-i	-ide
Locative	-s	-in
Comitative	-in	-iguin
Essive		-n

by three general patterns. First, case endings are invariant, as illustrated in Table 3.4, so the endings can be memorized and need not be determined for individual first-declension nouns. Second, the comitative singular and locative plural are always identical, so a speaker must encounter at most one of these two forms. The third and most fundamental pattern relates to stem alternations. Given that endings are invariant, solving the PCFP for an item reduces to the problem of determining the distribution of strong and weak stems. This task is made much easier by the fact that the cells of a first-declension paradigm divide into the same two "cohort sets" in the weakening and strengthening patterns. Set A contains the nominative and illative singular and essive, and Set B contains the remaining cells. In nouns of the weakening type, Set A is strong and Set B is weak; in nouns of the strengthening type, Set A is weak and Set B is strong.

A striking consequence of this symmetry is that **every** form of a first-declension noun is diagnostic. A strong form from Set A identifies a noun as belonging to the weakening type, and licenses the deduction that the remaining Set A forms are strong and the Set B forms are weak. Conversely, a weak form from Set A identifies a noun as belonging to the strengthening type, and licenses the deduction that the remaining Set A forms are weak and the Set B forms are strong. Any Set B form, whether strong or weak, is equally diagnostic. In sum, knowing the form of any one paradigm cell eliminates nearly all uncertainty about the forms that fill the other cells in a first declension paradigm. This implicational structure is completely symmetrical. Each form of a paradigm is equally informative, and the nominative and accusative singular forms that realize noun stems play no privileged role in distinguishing noun types.

A straightforward application of information-theoretic notions provides a natural means of measuring the implicational structure of the Saami system. To measure the uncertainty of forms in an inflectional paradigm *P*, we let *P* be a matrix whose dimensions are defined by features, and a paradigm cell *C* be a variable which takes as values the different realizations of the features associated with *C*. If the entropy of each cell of the Saami paradigm is 1 bit, and there are eleven cells in the paradigm, then if all cells were independent we would expect the overall entropy of the paradigm (that is, the joint entropy of all the cells) to be 11 bits. However, there are only two subdeclensions in Table 3.3, and if we again assume that each is equally likely, then the overall entropy of the paradigm is also 1 bit. This shows that there is a significant amount of shared information in the Saami paradigm. In fact, once you know the realization of one cell, you know the realization of every other cell: any one cell completely predicts the others. One can quantify the degree of prediction between these cells using entropy. The average uncertainty in one variable given the value another is the **conditional entropy** $H(Y|X)$. If $P(y|x)$ is the conditional probability that $Y = y$ given that $X = x$, then the conditional entropy $H(Y|X)$ is:

$$H(Y|X) = -\sum_{x \in X} P(x) \sum_{y \in Y} P(y|x) \log_2 P(y|x)$$

Conditional entropy can also be defined in terms of joint entropy:

$$H(Y|X) = H(X,Y) - H(X)$$

The smaller that $H(Y|X)$ is, the more predictable *Y* becomes on the basis of *X*, i.e., the less surprised one is that *Y* is selected. In the case where *X* completely determines *Y*, the conditional entropy $H(Y|X)$ is 0 bits: given the value of *X*, there is no question remaining as to what the value of *Y* is. On the other hand, 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*.

Given the paradigm in Table 3.3, we can calculate the conditional entropy of any one cell given any other cell. Let us take the nominative singular and the locative plural, which happen to belong to different cohort sets. Each cell has two possible realizations, and the entropy of each is 1 bit. To find the joint entropy, we look at the four possible combinations of realizations:

Nom Sg	Loc Pl	*P*
strong	strong	0.0
strong	weak	0.5
weak	strong	0.5
weak	weak	0.0

Once again, we have two equally likely possible outcomes, and the joint entropy is 1 bit. So, the conditional entropy is:

$$\begin{aligned} H(\textsc{loc.pl}|\textsc{nom.sg}) &= H(\textsc{nom.sg}, \textsc{loc.pl}) - H(\textsc{nom.sg}) \\ &= 1.0 - 1.0 \\ &= 0.0 \end{aligned}$$

That is, knowing the nominative singular realization for a particular noun completely determines the realization of the locative plural. One could repeat this calculation for any pair of cells in the paradigm and we would get the same result, as the Saami nominal inflection is a completely symmetric system.

In contrast, merely knowing one or both of the stem forms of a noun does not reduce uncertainty about whether a noun is of the weakening or strengthening type, because one must still know whether **which cell** the stem realizes. Knowing that the noun BIHTTÁ in Table 3.3 has the strong stem *bihttá* and the weak stem *bihtá* does not identify the subtype of this noun unless one knows which stem underlies which cohort set. Knowing that BASTE has the strong stem *bastte* and the weak stem *baste* is similarly uninformative. Hence, the type of these nouns cannot be determined from their stem inventories but only from the distribution of stems in the inflectional paradigms of the nouns.

3.3.2 *Finnish*

The Finnish subparadigm in Table 3.5 illustrates a more typical pattern, in which different **combinations** of cells are diagnostic of declension class membership.[9] Although individual forms may be indeterminate with respect to class membership, particular combinations of forms in Table 3.5, varying from class to class, reduce the uncertainly of class assignment. Consider forms *laseissa, nalleissa* and *kirjeissa*, which realize the inessive plural in the paradigms of the nouns of LASI, NALLE, and KIRJE. None of these forms alone reliably predicts the corresponding nominative singular forms. But collectively

[9] The numbers in Table 3.5 refer to the declension classes in Pihel and Pikamäe (1999).

TABLE 3.5 Finnish *i*-stem and *e*-stem nouns (Buchholz 2004)

Nom Sg	Gen Sg	Part Sg	Part Pl	Iness Pl	
ovi	oven	ovea	ovia	ovissa	'door' (8)
kieli	kielen	kieltä	kieliä	kielissä	'language' (32)
vesi	veden	vettä	vesiä	vesissä	'water' (10)
lasi	lasin	lasia	laseja	laseissa	'glass' (4)
nalle	nallen	nallea	nalleja	nalleissa	'teddy' (9)
kirje	kirjeen	kirjettä	kirjeitä	kirjeissä	'letter' (78)

they provide information that the appropriate class is restricted to 4, 9, or 78, but not 8, given that the inessive plural in class 8 is *ovissa*, not *oveissa*. Certain cells among these classes resolve class assignment more reliably than others. For example, *kirjeitä*, the partitive plural of KIRJE, appears unique among the forms in the partitive plural column and, therefore, is serviceable as a diagnostic cell for membership in class 78. This becomes particularly clear when we compare this form with the partitive plural forms *laseja* and *nalleja*: even in conjunction with the previously mentioned inessive plurals, these forms do not resolve class assignment between 4 and 9. This is accomplished, however, by comparing the partitive singular forms, *lasia* and *nallea*, or several other contrasts that would serve just as well.

These class-specific sets are reminiscent of the notion of **dynamic** principal parts, which Finkel and Stump (this volume) contrast with what they term "static" and "adaptive" inventories. In fact, there are many equally good alternative sets of principal parts for Finnish, and many more solutions that are almost as good. We speculate that this is a recurrent feature of complex morphological systems (reminiscent of resilience in biological systems). Even though there may be a few very hard cases or true irregulars, in general most cells in the paradigm of most words are of value in predicting the form of most other cells.

As the traditional principal part inventories in Table 3.5 show, the information that facilitates paradigm cell filling in Finnish is not localized in a single form or even in a class-independent set of forms. Instead, forms of an item are partitioned into cohort sets or "subparadigms" that share "recurrent partials." One pair of subparadigms in Finnish declensions are distinguished by what are conventionally termed the "basic form" and the "inflectional stem" of an item. A typical pattern is illustrated by the paradigm of OVI 'door', in which the basic form *ovi* realizes the nominative singular and underlies the partitive and inessive plurals, and the inflectional stem *ove* underlies the genitive and partitive singular forms. As in Saami, the organization of cells into subparadigms identifies the form of other declensional cohorts, while variation in the structure of subparadigms across items facilitates the identification of declension classes.

Given this overview of the patterns in Table 3.5, we now outline how to calculate the joint and conditional entropy of the corresponding paradigm cells. Let us first consider how many distinct realizations of the genitive singular are exhibited in Table 3.5. From a traditional perspective, there is exactly one affixal realization, given that "[t]he genitive singular ending is always -*n*, which is added to the inflectional stem" (Karlsson 1999: 91). However, this description already presupposes knowledge of the inflectional stem, which is precisely the type of information that a speaker may need to deduce in order to solve the PCFP. To avoid presupposing information about the organization of Finnish declensions, it is useful to adopt more structurally agnostic descriptions in terms of a "base", which underlies the basic form (and, usually, the inflectional stem), and an "ending" (which may include the theme vowel of the inflectional stem).[10] On this type of description, the six inflectional classes in Table 3.5 exhibit four distinct realizations. In classes 8, 32, and 9, the genitive singular ends in -*en*. In class 10, it ends in -*en* and the base exhibits a change in the stem consonant. In class 9, it ends in -*in*, and in class 78 it ends in -*een*. If we assume that each of the six declensions has a probability of $\frac{1}{6}$, then the entropy H(GEN.SG) is:

$$H(\text{gen.sg}) = -\left(\frac{3}{6}\log_2\frac{3}{6} + \frac{1}{6}\log_2\frac{1}{6} + \frac{1}{6}\log_2\frac{1}{6} + \frac{1}{6}\log_2\frac{1}{6}\right)$$
$$= 1.792$$

Repeating this calculation for each of the cells in the paradigm, we get: The **expected** entropy $E[H]$ is the average across all cells. Producing a randomly

	Nom Sg	Gen Sg	Part Sg	Part Pl	Ines Pl	$E[H]$
H	0.918	1.792	2.252	1.459	1.000	1.484

chosen cell of the paradigm of a randomly chosen lexeme (assuming that the declensions are equally likely) requires on average 1.484 bits of information.

Given the paradigms in Table 3.5, we can also calculate the pairwise conditional entropy. Suppose we know that the NOM.SG of a particular lexeme ends in -*i*. What is the genitive singular? Our information about the NOM.SG rules out classes 9 and 78, so we are left choosing among the remaining four classes with three different GEN.SG realizations. Given this information, the uncertainty in the GEN.SG becomes:

[10] This type of pretheoretical description is found particularly in pedagogical grammars and descriptions. For example, noun classes 16–22 in Oinas (2008: 57f.) distinguish *i*- and *e*-stem nouns in terms of the surface variation in their genitive singular forms.

$$H(\text{GEN.SG}|\text{NOM.SG} = \text{-}i) = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$$
$$= 1.5$$

In other words, knowing that the NOM.SG ends in *-i* gives us 1.793 − 1.5 = 0.292 bits of information about the form of the GEN.SG. And, if instead we know that the NOM.SG of a particular lexeme ends in *-e*, then we must choose between two declensions with two GEN.SG realizations, and the entropy is:

$$H(\text{GEN.SG}|\text{NOM.SG} = \text{-}e) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$$
$$= 1$$

Assuming again that all declensions are equally likely, the probability that the NOM.SG of a particular lexeme actually ends in *-i* is $\frac{4}{6}$, and the probability that it ends in *-e* is $\frac{2}{6}$. So, on average, the uncertainty in the GEN.SG realization of a lexeme given we know that lexeme's NOM.SG realization will be:

$$H(\text{GEN.SG}|\text{NOM.SG}) = \frac{4}{6} \times 1.5 + \frac{2}{6} \times 1.0$$
$$= 1.333$$

In other words, the NOM.SG gives us, on average, 1.793 – 1.333 = 0.46 bits of information about the GEN.SG. Table 3.6 gives the pairwise conditional entropy of a column given a row. That is, e.g., H(NOM.SG|INESS.PL) is 0.541 bits.

The row expectation E[row] is the average conditional entropy of a column given a particular row. This is a measure of the **predictiveness** of a form. By this measure, the partitive singular is the most predictive form: if we know the partitive singular realization for a lexeme and want to produce on other paradigm cells chosen at random, we will require only 0.250 bits of additional information on average. In contrast, given the nominative singular, we would

TABLE 3.6 Conditional entropy H(col|row) of Finnish *i*-stem and *e*-stem nouns

	Nom Sg	Gen Sg	Part Sg	Part PI	Ines PI	E[row]
Nom Sg	—	1.333	1.667	0.874	0.541	1.104
Gen Sg	0.459	—	0.459	0.459	0.459	0.459
Part Sg	0.333	0.000	—	0.333	0.333	0.250
Part PI	0.333	0.792	1.126	—	0.000	0.563
Ines PI	0.459	1.252	1.585	0.459	—	0.939
E[col]	0.396	0.844	1.209	0.531	0.333	0.663

need an addition 1.104 bits of information on average. The column expectation *E*[col] is the average uncertainty given a row remaining in a particular column.

In contrast to the row expectations, this is a measure of the **predictedness** of a form. By this measure, the inessive plural is the most predicted form: if we want to produce the inessive plural for a lexeme and know some randomly selected other form, we will require on average another 0.333 bits of information.

One cannot of course draw any general conclusions about the implicational structure of Finnish declensions from the calculations in Table 3.6, given that they are based on a small subset of patterns, and that they assume that all classes and variants are equiprobable. Nevertheless, it should be clear that the method applied to this restricted data set scales up, as the description becomes more comprehensive through the addition of further patterns and as it becomes more accurate through the addition of information about type and token frequency.

3.3.3 *Tundra Nenets*

The present section now extends the approach outlined above in order to provide a preliminary case study of nominal inflection in Tundra Nenets (Samoyed branch of Uralic). The basic question is this: Given any Tundra Nenets inflected nominal word form, what are the remaining 209 forms of this lexeme for the allowable morphosyntactic feature property combinations CASE: {nom, acc, gen, dat, loc, abl, pro}, NUMBER: {singular, dual, plural}, POSSESSOR: {3 persons × 3 numbers}? The problem can be schematized as in (5a) and (5b). Specifically, given exposure to a stimulus such as that in (5a), the nominal *nganu′′mana* 'boat (plural prosecutive)', what leads to the inference that its nominative singular form is the target *ngano*? In contrast, if confronted with the plural prosecutive of the nominal *wín′′oqmana* 'tundra (plural prosecutive)', what leads to the inference that its nominative singular is the target *wí*?

(5)	a.	Stimulus:	Target	vs	b.	Stimulus	Target
		nganu′′mana	*ngano*			*wíngo′′mana*	*wí*
		boat.PL.PROS	boat.SG.NOM			tundra.PL.PROS	tundra.SG.NOM

In line with the hypotheses set out in the previous section, we must identify the patterns of interpredictability for a subset of Tundra Nenets nominal declensions within and across subparadigms. This entails stating the principles of arrangement within and across stem types. For the absolute declension (i.e., nonpossessive, nonpredicative nominals), lexical categories are divisible into the gross stem-type classification in Table 3.7 (again ignoring the role of syllabicity;

Table 3.7 Tundra Nenets nominal types (Salminen 1997, 1998)

Type 1 (T1):	stem ends in C (other than a glottal stop) or V;
Type 2 (T2):	subtype 1: stem ends in nasalizing/voicing glottal (′)
	subtype 2: stem ends in non-nasalizing/devoicing glottal (″)

see Salminen (1997, 1998) for a detailed exposition of types).[11] For simplicity, we demonstrate the basic pattern with an example of Type I in Table 3.8.

Examination of Table 3.8 yields a basic observation: the nominal paradigms for all stem classes are partitioned into subparadigms, each of which is defined by the presence of a characteristic and recurring stem (*ngano, nganu*, or *nganoxo*). In what follows we will refer to these forms as recurrent partials and the sets in which they recur as coalitions or alliances (or cohorts) of forms. This brings out the following generalization about Tundra Nenets absolute nominal paradigms:

Subparadigms are domains of interpredictability among alliances of word forms, rather than sets of forms derived from a single base.[12]

An approach based on recurrent partials, and patterns of relatedness among forms, develops the approach in Bochner (1993), in which no form need serve as a privileged base form among different surface expression of a lexeme.

Table 3.8 Type I: Polysyllabic vowel stem: *ngano* 'boat'

	Singular	Plural	Dual
Nominative	*ngano*	*ngano″*	*nganoxo′*
Accusative	*nganom′*	*nganu*	*nganoxo′*
Genitive	*ngano′*	*nganu″*	*nganoxo′*
Dative-Directional	*nganon′*	*nganoxo″*	*nganoxo′ nya′*
Locative-Instrumental	*nganoxona*	*nganoxo″ na*	*nganoxo′ nyana*
Ablative	*nganoxod*	*nganoxot*	*nganoxo′ nyad*
Prolative	*nganowna*	*nganu″ mana*	*nganoxo′ nyamna*

[11] There are phonological properties associated with particular glottal-final stems (as in Saami and Finnish) that decrease the uncertainty of predicting class assignment and related forms of words within the class. For example, the occurrence of a specific allomorph, e.g., *wíngana* (where *-gana* is part of a family allomorphs such as *-xana* and *-kana*) leads to the inference that this word belongs to the class of stem-final nasalizing glottals. In this way, surface allomorphy can be used as a diagnostic clue for guiding paradigm-based inferences.

[12] As expected from positing that patterns of inflected forms exist, there is a need to access certain of them for purposes of derivational relatedness in Tundra Nenets. In particular, there are at least two verbal derivation operations built upon the form used to express genitive plural nominals. See Kupryanova (1985: 139).

Regardless of whether a stem exists as an independent word, all these systems share the property that they have clusters of related forms where it is at least somewhat arbitrary to take any one form as basic. This is what I take to be defining characteristic of a paradigm. Thus, we need a way to relate to the various members of paradigm directly to each other without singling out any one of them as a base for the others. (Bochner 1993: 122)

On this type of analysis, alliances of word forms share recurrent partials, but the elements in such alliances need not be thought of as bearing derivational or "constructive" relations (in the sense of J. P. Blevins 2006*b*) to one another, let alone to a single isolable base form. The relations among members of subparadigms are symmetrical, since there is no one form that serves as the base from which the others are derived.[13] This organization is depicted in Figure 3.1, which partitions the Tundra Nenets nominal declension into three alliances of forms. Each form in a subparadigm provides information about other forms in the same subparadigm. The members of a subparadigm share partials, thereby making an alliance a system of interpredictability among related word forms.[14]

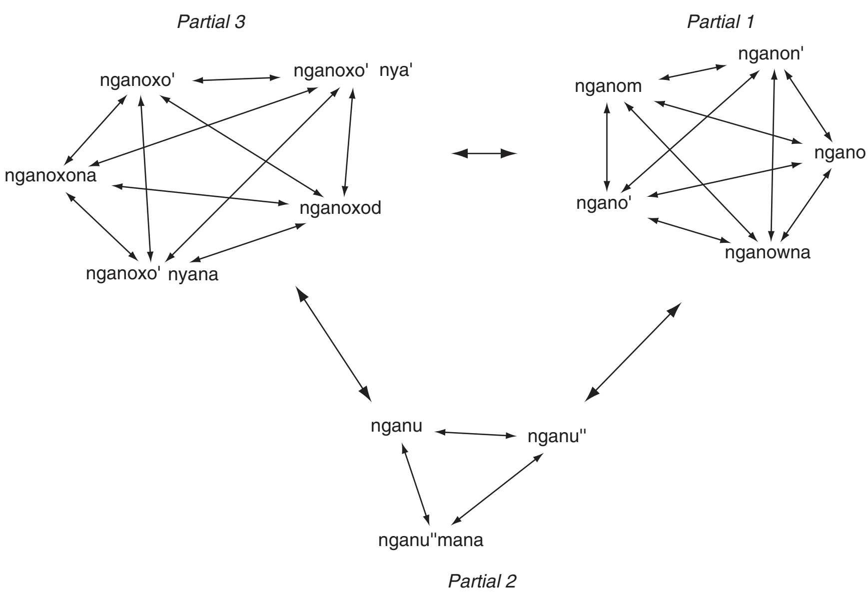

Figure 3.1 Symmetric paradigm organization

[13] However, the lack of a single privileged base does not entail that there cannot be multiple subparadigms in which a particular recurring form (a partial) serves a pivotal role.

[14] This is compatible with Albright's observation that "when we look at larger paradigms . . . it often appears that we need local bases for each sub-paradigm (something like the traditional idea of principal parts, or multiple stems)" (Albright 2002*a*: 118).

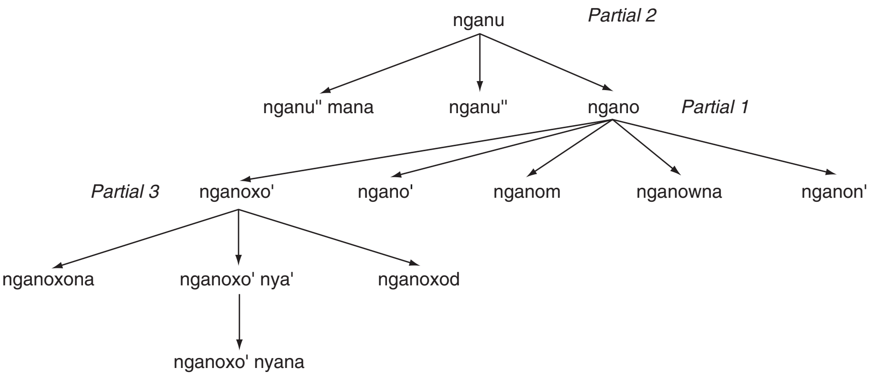

Figure 3.2 Asymmetric paradigm organization

In contrast, derivational or constructive relations based on a single form are asymmetric in assuming that some specific form is predictive of the other forms. An asymmetric structure, organized in in terms of local bases, is depicted in Figure 3.2. In contrast to Figure 3.1, each subparadigm contains a base from which the rest of the forms in it are derived. There is no notion of interpredictability of the sort manifest in Figure 3.1: the base gives information about derived forms, but the derived forms need not give information about a base.

3.3.3.1 *Implications across subparadigms* The strategy we have chosen to demonstrate the utility of symmetric organization is to focus on the most challenging and problematic instance of relatedness between two word forms within Tundra Nenets nominal paradigms, specifically the nom.sg and acc.pl. The logic of this task is straightforward: if we can identify a direction with reliably low conditional entropy, i.e., most predictive, between the two least transparently related word forms, then there is reason to believe that asymmetric derivation may be viable. In other words, one could hypothesize that knowing e.g., nom.sg, would suffice to predict the acc.pl across all classes, either directly, or by identifying a common base that underlies both forms. In contrast, the symmetric proposal is compatible with a situation in which there is no single reliably predictive form, but that classes are organized into patterns of interpredictability within alliances of forms.

Consider the pairs of nom.sg and acc.pl forms in Table 3.9. A comparison of the forms in the columns reveals that there is indeterminacy or uncertainty with respect to predictability in both directions. For example, while the acc.pl of 'boat' and 'harnessed deer' both end in the vowel -*u*, their nom.sg forms

TABLE 3.9 Tundra Nenets inflected nominals

Nom Sg	Acc PI	
ngano	nganu	'boat'
lyabtu	lyabtu	'harnessed deer'
ngum	nguwo	'grass'
xa	xawo	'ear'
nyum	nyubye	'name'
yí	yíbye	'wit'
myir	myirye	'ware'
wí´	wíngo	'tundra'
we´	weno	'dog'
nguda	ngudyi	'hand'
xoba	xob	'fur'
sawənye	sawənyi	'magpie'
tyírtya	tyírtya	'bird'

end in -*o* and -*u* respectively. Likewise, while the NOM.SG of 'boat' ends in -*o*, the ACC.PL of 'grass' ends in -*o* and its NOM.SG ends in the consonant -*m*.

The basic question is, given exposure to one form, how well can one predict the other? This is just the PCEP relativized to Tundra Nenets. In the following preliminary study, we use data from a corpus of 4,334 nominals. These are extracted from Salminen's compilation of 16,403 entries, which is based on Tereshchenko's Nenets-Russian dictionary (1965/2003). The compilation specifies meaning, frequency, as well as the stem-class assignment. We explore the relative predictiveness of NOM.SG and ACC.PL, with the following query in mind: which of these forms, if either, is more useful for predicting the other? The first calculation maintains the idealization adopted in the analyses of Saami and Finnish and assumes that all declension classes are equally likely. We start by identifying 24 different types of nominative singulars. The entropy of this distribution is $H(\text{NOM.SG}) = 4.173$ bits. There are likewise 29 different types of accusative plurals, and their entropy is $H(\text{ACC.PL}) = 4.693$ bits. Taken together, there are 43 nominal 'declensions' represented in the compilation (each declension being a combination of a NOM.SG realization and an ACC.PL realization), and the joint entropy of the two forms is $\log_2 43 = 5.426$ bits.

These calculations assume (as in the case of Saami and Finnish) that all declensions are equally likely. However, it is clear from the compilation that all declensions are **not** equally likely. In fact, the distribution of type frequencies across declensions is highly skewed: the five most frequent declensions account for more than half of the noun lexemes (see Figure 3.3 for the complete distribution). Taking the type frequencies of declensions into account, we now find that the entropy associated with each individual form is H(NOM.SG) = 3.224 bits and H(*acc.pl*) = 3.375 bits. The true joint entropy H(NOM.SG, ACC.PL) is 3.905 bits, a level of uncertainty equivalent to 15 equiprobable declensions.

Having quantified the degree of uncertainty in the choice of NOM.SG and ACC.PL types individually, we can now calculate predictability of one realization given the other, using conditional entropy $H(Y|X)$. Consider first the task of predicting the ACC.PL form from the NOM.SG. We can evaluate the difficulty of this prediction using the conditional entropy H(ACC.PL|NOM.SG), the uncertainty in the ACC.PL given the NOM.SG. Out of the $24 \times 29 = 696$ possible pairings of NOM.SG and ACC.PL types, 43 are actually attested in the lexicon. In some cases, knowing the NOM.SG of a word uniquely identifies its ACC.PL, e.g. a word ending in -*ye* in the NOM.SG always has an ACC.PL in -*yi*. For such words, once we know the NOM.SG there is no uncertainty in the ACC.PL and the conditional entropy H(ACC.PL|-*ye*) = 0 bits. In other cases, however, knowing the NOM.SG narrows down the choices for the ACC.PL but does not uniquely identify it. For example, polysyllabic words whose NOM.SG ends in -*ya* might have an accusative plural in -∅, -*yi*, or -*e*. Furthermore, of the 289 polysyllabic lexemes with a NOM.SG in -*ya*, 268 have an ACC.PL in -*yi*, 19 in -∅, and only 2 in -*e*. So, the entropy is:

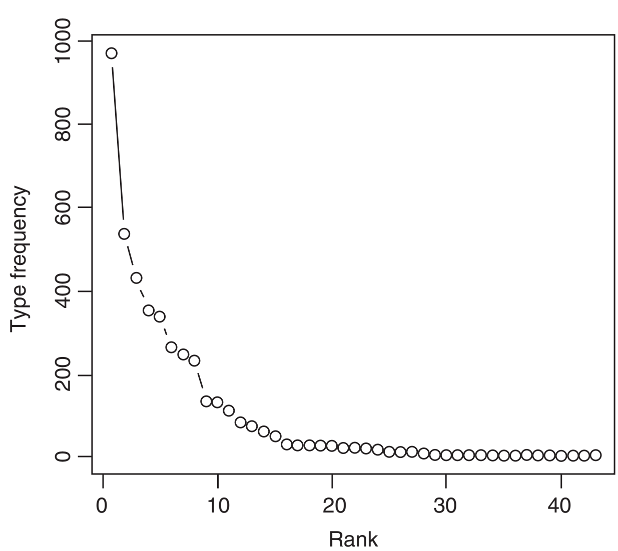

FIGURE 3.3 Type frequencies of Tundra Nenets nominal declensions, by rank

$$H(\text{ACC.PL}|\text{NOM.SG} = \textit{-ya}) = -\left(\frac{268}{289}\log_2\frac{268}{289} + \frac{19}{289}\log_2\frac{19}{289} + \frac{2}{289}\log_2\frac{2}{289}\right)$$
$$= 0.410 \text{ bits}$$

Averaging across the whole (sample) lexicon, the uncertainty in the ACC.PL given the NOM.SG is $H(\text{ACC.PL}|\text{NOM.SG}) = 0.681$ bits. In other words, the NOM.SG "predicts" all but 0.681 of the 3.375 bits of uncertainty previously calculated for the ACC.PL. Now, if we switch directions, going from ACC.PL to NOM.SG, it turns out that the conditional entropy $H(\text{NOM.SG}|\text{ACC.PL}) = 0.530$. In other words, the ACC.PL "predicts" all but 0.530 of the 3.224 bits in the NOM.SG. Since the conditional entropy is closer to 0 in the latter than in the former, the ACC.PL appears to be more helpful for predicting the NOM.SG than vice versa, but only by a slim margin. More importantly, neither conditional entropy is 0 bits or close to it, meaning neither form is especially useful for predicting the other.

Hence, there is no principled grounds for hypothesizing that one form or the other serves as (or even identifies) a single privileged base. Either choice would still leave a large inventory of irregular pairings to be memorized by the language learner. This arbitrary choice is avoided on a symmetric account, where there is no need to suppose that some forms are reliably predictable from others. Instead, a symmetrical proposal posits alliances which cohere into coalitions of interpredictable forms and which together partition the entire paradigm. We do not expect forms that take part in different alliances to be mutually predictive, so the fact that knowledge of a member of one alliance does not reliably reduce uncertainty about a member of another is not surprising.

More positively, the utility of alliances becomes clearer if one considers the distribution of Tundra Nenets forms. Although the NOM.SG and ACC.PL are equally unsuitable as single bases, the NOM.SG will still make a more prominent contribution to defining the implicational structure of a paradigm, given that speakers are far more likely to encounter the NOM.SG form of a noun than the ACC.PL form. The distributional difference between these forms is reflected in the frequency counts in Table 3.10, representing the 12,152 noun tokens in Salminen's sample sentence corpus. The NOM.SG represents 33.8 percent of the tokens, while the ACC.PL represents only 2.7 percent. Speakers cannot just assume that the most frequent form is the most useful for solving the PCFP, given that the NOM.SG is not even a reliable predictor of the ACC.PL. The ACC.PL itself is an even less suitable candidate. Even if the predictive value of the ACC.PL made it potentially useful as a base, the attested frequencies suggest that speakers would have a low likelihood of encountering this form for any

TABLE 3.10 Word-form frequencies in Tundra Nenets

	Singular	Plural	Dual
Nominative	4,117	770	7
Accusative	1,077	355	6
Genitive	3,002	376	5
Dative-Directional	762	89	0
Locative-Instrumental	724	108	0
Ablative	291	50	0
Prolative	372	41	0

given item. The situation is worse yet for forms such as the direct case dual forms, which account for 0.1 percent of the tokens. In fact, no individual word form (other than the NOM.SG and the GEN.SG) occurs with high enough frequency to be a reliable source of information about a word's inflectional class. This makes Tundra Nenets a challenging language from a "single base" point of view, as speakers cannot be sure of encountering the diagnostic forms necessary to identify a word's inflection class.

However, the issue takes on a different complexion when we look at forms in terms of alliances, organized around the Partials 1, 2 and 3 in Figure 3.1. Although the ACC.PL is a relatively low-frequency form, it is predictable from other forms that it is transparently related to. For example, the GEN.PL adds a final glottal stop to the ACC.PL, as illustrated by the relation between *nganu*, the ACC.PL form of 'boat', and the corresponding GEN.PL *nganu´´*. Hence, while there is a low likelihood of encountering the ACC.PL, there is a much higher likelihood of encountering the **partial** associated with ACC.PL (from which the ACC.PL can be defined), if paradigms are organized into alliances of interpredictable forms that "pool" the frequency of individual forms. The effect of this structure is shown by the contrast between the form frequencies in Figure 3.1 and the totals in Table 3.11, which sum the token frequencies of all absolute and possessive forms.

The organization of forms into subparadigms thus serves two related functions. On the one hand, high-frequency forms such as the NOM.SG or GEN.SG identify the shape of lower-frequency members of the same alliance, such as the prolative singular. On the other hand, "pooling" the frequencies of the members of each alliance allows Partial 2, and the forms based on this partial, to be identified either by the ACC.PL and the GEN.PL, while Partial 3, and forms based on it, can be identified by the locative-instrumental forms or by the ablative singular. By relying on alliances of related forms within subparadigms, speakers may gain reliable cues about the shape of even very low-frequency word forms.

Significantly, accounts that assume an asymmetrical relation between a privileged base and derived forms have no obvious analogue to alliances of mutually reinforcing forms. On such asymmetrical approaches, the patterns within subparadigms appear epiphenomenal, not, as suggested here, as central to the organization of the declensional system and critical to the solution of the PCFP.

3.3.3.2 *Summary* The preceding sections suggest that neither the NOM.SG nor ACC.PL form can serve reliably as the single base from which the other is predicted. Yet the fact that neither form is fully predictive does not mean that they are uninformative. Instead, the association of forms with subparadigms allows speakers to exploit the fact that partials appear with much higher frequency than any given word-form. Hence, there is no need to encounter a privileged member of an alliance in order to predict allied forms. What is important is just that each alliance contain at least some high-frequency forms and that the the aggregate frequency of partials within the alliance is high enough to be useful. In this way, the organization of the Nenets declensional system makes available many of the basic ingredients for a solution to the paradigm cell filling problem.

3.4 Conclusions

We conclude by returning to the questions in (6), repeated from (1), which concern issues raised by traditional solutions to the Paradigm Cell Filling Problem.

(6) a. What is the structure of units that license implicative relations?
b. How are units organized into larger structures within a system?
c. How can one measure implicative relations between these units?
d. How might the implicative organization of a system contribute to licensing inferences that solve the paradigm cell filling problem?
e. How does this organization, and the surface inferences it licenses, contribute to the robustness and learnability of complex systems?

This chapter has focused primarily on questions (6a), (6b), and (6c). The central hypothesis has been that words are organized into paradigms and that information-theoretic measures provide an insightful measure of relatedness among members of declension classes. In fact, distinctive patterns of relatedness clearly enter into what it means to be a declension class, with some forms or combinations of forms being more diagnostic of class membership than others. Once conditional entropies for families of forms are identified, they can be used, along the lines we have suggested, to solve the Paradigm Cell

Filling Problem. Individual forms or alliances of forms serve as cues for simplifying the assignment of class membership for novel words on the basis of the analogies provided by the patterns of known words. It is worth emphasizing that the answers to questions (6a), (6b), and (6c) presuppose access to (patterns of) surface word forms. There are, accordingly, several theoretical consequences associated with our results.

First, there must be more to morphological analysis and morphological theory than the distillation of rules or patterns for the composition of individual word forms. In focusing exclusively interest on the syntagmatic dimension of morphological analysis, the post-Bloomfieldian tradition has been led to adopt questionable claims about the nature of the grammatical system and the mental lexicon. Work within this tradition has assumed that morphological analysis consists of identifying morphemes and stating rules for describe morpheme combinations. Larger structures such as words and paradigms tend to be treated as derivative or even as epiphenomenal. The emphasis on identifying minimal units has also fostered the *a priori* belief that the lexicon consists entirely of minimal elements, and, in particular, that productive and regular word forms are not part of the mental lexicon of a speaker, on the grounds that such forms would be "redundant" if they could be constructed from available morphemes and combinatoric rules. Yet a range of psycholinguistic studies has shown that the processing of a given word may be influenced (whether facilitated or inhibited) by other related forms in a way that suggests that the related words are available as elements of a speaker's mental lexicon (Baayen *et al.* 1997; Schreuder & Baayen 1997; Hay 2001; de Jong 2002; Moscoso del Prado Martín 2003). Another group of studies provide evidence for various types of paradigm-based organization (Baayen & Moscoso del Prado Martín 2005; Milin *et al.* 2009).

The traditional word and paradigm assumptions adopted here appear to be more compatible with these results than the post-Bloomfieldian assumptions that still guide modern generative accounts. In order to unify these perspectives, one might take them to adopt have complementary foci, with WP approaches focusing on whole words and their organization into paradigms, and morphemic accounts focusing on the internal structure and construction of word forms. We suggest that this is misleading. For the languages discussed above and others of comparable complexity, the answer to question (6a) must appeal to the whole words and larger paradigmatic structures recognized in WP approaches. There is little evidence that syntagmatic approaches have any means of characterizing the role that whole words play in morphology, let alone the place of larger paradigmatic structures. In contrast, a WP approach is largely agnostic about the internal structure of complex words. A WP approach is compatible with an agglutinative **morphotactic** analysis, in cases where such an analysis is motivated. But a WP account is also able to characterize the extraordinary variety of strategies for the creation of complex word forms attested cross-linguistically, without reducing them to an underlying basic structure. In order to arrive at a general answer to question (6a), we suggested that complex words are recombinant gestalts. On this pattern-based view, agglutination is just a particularly simple pattern. Finally, with respect to question (6e), we suggest that it is the very pattern-based nature of morphology – both at the level of individual (types of words) and in their organization into paradigms – that makes even highly complex morphological systems learnable and, by hypothesis, guides the development, maintenance, and change of these systems.