of agreement. Rather, the optionality is in how the relevant paradigm is represented. While some verbs are attested in which the inflectional morphology directly represents the features of the arguments, or tense category, we have seen that it is quite common for verbs in Kanum to extend the range of one cell of a paradigm into another. While there are constraints on the kind of extensions found, the appearance of such an extension in a verb's paradigm is not predictable.

4.8 The middle ground between productivity and suppletion

The preceding discussion has raised the questions of predictability, productivity, and complexity (see also Bauer 2004). The presence of rules of referral in any verb's paradigm is lexically unpredictable, but is highly productive. It does not involve the use of novel or suppletive forms, but rather involves the extension (or, in some cases, transfer) of a form from one cell of a paradigm to another. In Kanum we have seen that takeovers operate most strongly for tense, but that object agreement is most strongly implicated in the more complicated (in the sense of unpredictable) verbal paradigms. Whether these patterns can be found in other languages or not remains to be seen.

5

Morphological complexity à la Oneida

JEAN-PIERRE KOENIG AND KARIN MICHELSON

This chapter is about one particularly rich part of the verbal inflection of Oneida, a polysynthetic Iroquoian language. Morphological referencing of event participants in Oneida is achieved via a system of fifty-eight pronominal prefixes that are an obligatory part of the inflection of verbs. The sheer number of prefixes, and the relations between them, afford us the opportunity to ask what we believe is a unique set of questions about morphological complexity.

Morphological complexity differs from syntactic complexity in significant ways. Issues of morphosyntactic complexity have been mostly approached from two perspectives: computational complexity, which results from the concatenation of components in a construction (a phrase, or a sentence), and algorithmic complexity, which is due to processing sentences in real time. Whether complexity is evaluated at the computational or algorithmic level, the kind of complexity that has interested syntacticians is syntagmatic, that is, it arises from the fact that the formatives of a sentence are linearly sequenced. In contrast, morphological complexity is paradigmatic in nature. Thus issues of morphological complexity stem, for example, from the need to select an affix (in the case of affixal morphology) among many possible choices and the need to segment words into their formatives (or at least analyse words for purposes of classifying inflected words into the right paradigm; see Blevins 2013 for discussion). Of course, that morphological complexity differs from syntactic complexity is not novel, and some of the issues that have been addressed specifically about inflectional systems are the limits on inflectional classes (e.g. Carstairs 1983, Carstairs-McCarthy 1994, and discussion in Blevins 2004), optimal principal part systems (e.g. Finkel and Stump 2007, Finkel and Stump 2013), and entropy (e.g. Ackerman et al. 2009). Our focus here is to show that the two aspects of language production and comprehension that were mentioned as specific to morphology—the paradigmatic notions of selection and
segmentation—can lead to enormous complexity even when one considers just a single block of inflectional realizations (or, in traditional terms, a single position class). 1

Oneida (Iroquoian) verbs include bound, obligatory prefixes that reference participants in the situations described by verbs. Thus, the verb form in (1) includes the prefix luwa-, which indicates that the described situation includes a third person feminine singular, third person indefinite, or third person dual or plural agent acting on (represented by '>') a third person masculine singular patient, while the verb form in (2) includes the prefix shukwa-, which indicates that the described situation includes a third person masculine singular agent acting on a first person plural patient. 2

(1) luwa-hlo li-he
  3>3M.SG-tell-HAB
  'she, someone, or they tell him'

(2) shukwa-hlo li-he
  3M.SG>1PL-tell-HAB
  'he tells us'

The prefixes illustrated in (1) and (2)—traditionally labelled pronominal prefixes—occur in a single position class (i.e. in incremental or realizational terms, are the output of a single rule block). There are fifty-eight of these prefixes and ever since Lounsbury’s (1953) seminal work, the complexity of Oneida pronominal prefixes has been considered a hallmark of Iroquoian languages (and a challenge for second language learners; see Abrams 2006). Oneida pronominal prefixes, then, provide a rather unique case of paradigmatic complexity.

The bulk of our chapter is devoted to identifying the formal dimensions along which the Oneida pronominal prefix system may qualify as paradigmatically complex and we come back to the nature of paradigmatic complexity in the conclusion. Section 5.1 briefly describes Oneida pronominal prefixes. Section 5.2 identifies parameters of paradigmatic complexity. Sections 5.3, 5.4, and 5.5 focus on how Oneida pronominal prefixes stack up on three of these parameters (size of the space of morphological distinctions to be marked, semantic ambiguity, and directness). Section 5.6 discusses one aspect of paradigmatic complexity specific to Oneida pronominal prefixes that, to our knowledge, has not been discussed in the literature, and that is the possible misidentification of pronominal prefixes in verb forms. Section 5.7 concludes the chapter.

5.1 A brief description of Oneida pronominal prefixes

The Oneida verb has an elaborate internal structure. Stems can be complex, derived via prefixation, suffixation, and noun incorporation. In addition to the obligatory pronominal prefixes, verb forms must have an aspect suffix or an imperative ending (often ‘zero’). The aspectual categories are habitual (basically imperfective), punctual (perfective), and stative (having stative, perfect, or progressive meaning depending on the verb). Verbs in the punctual aspect always occur with one of three modal prefixes, the factual, future, or optative (usually the factual in this chapter). In addition to a modal prefix, verb forms can have one or more of eight other prepronominal prefixes.

An example of a typical verb in Oneida is given in (3).

(3) s-a-huwa-akt-a-hu-e
    REP-FACT-3>3M.DP-go.to.a.point-JN-CAUS-PNC
    ‘they pushed them back, they made them retreat’

Pronominal prefixes are portmanteau-like. Although it is often possible to associate parts of the prefixes with some features (e.g. ti or wa with plurality), it is widely accepted that a single prefix references one or two participants as a whole. Thus, although one may associate the initial l with masculine in the prefix luwa- ‘3>3M.SG’, one cannot segment the prefix into two subparts, each referencing a distinct participant in the described situation. More generally, even parts of prefixes most easily associated with particular attribute-value pairs do not allow a segmentation into proto-agent and proto-patient parts. Consider the prefixes li- (referring to a 1SG proto-agent acting on a 3M.SG proto-patient) and lak- (referring to a 3M.SG proto-agent acting on a 1SG proto-patient). One can recognize /l/ both as marking masculine gender, but it marks the masculine gender of the proto-patient in the first case and the masculine gender of the proto-agent in the second case.

Semantic categories distinguished by the pronominal prefixes are person (first, second, third, plus an inclusive/exclusive distinction), number (singular, dual, plural), and gender (masculine, feminine, feminine-zoic, neuter). In the singular, feminine-zoic gender is used for some female persons (see Abbott 1984) and for animals that are not personified and marked with the masculine. It is also used for all female

---

1 See Hankamer (1989) on Turkic, Fortescue (1980) on Greelandic Eskimo, as well as Anderson (this volume), for cases where morphology is syntagmatically relatively complex.
2 In the Oneida examples A is a mid, central, nasalized vowel, and U is a high or mid-to-high, back, nasalized vowel. A raised period represents vowel length. Voicing is not contrastive. Abbreviations used in the volume), for cases where morphology is syntagmatically relatively complex.

---

See Hankamer (1989) on Turkish, Fortescue (1980) on Greelandic Eskinio, as well as Anderson (this volume), for cases where morphology is syntagmatically relatively complex.
persons in the dual and plural, as the feminine gender is restricted to the singular. Note that neuter gender is a semantic category only (as explained later). In addition, there is a third person indefinite ('indefinite' in Lounsbury 1953, or 'nonspecific' in Chafe 1977) translated as 'one, people, they, someone'. Because of the number of properties the prefixes can reference and because, as we will elaborate, prefixes mark up to two semantic arguments, the number of prefixes for this inflectional slot is quite large, fifty-eight in total. The fifty-eight prefixes are given in Table 5.1, based on Table 6 in Lounsbury 1953; the prefixes are numbered as in Lounsbury, and later on, when we refer to specific prefixes, we often identify the prefix with this number, preceded by 'I' (for Lounsbury).

The prefixes fall into two categories. 'Transitive' prefixes mark two animate arguments, as in the example in (4), repeated from (2). In Table 5.1, the properties of the proto-agent that are marked by transitive prefixes are given in the leftmost column and properties of the proto-patient are given in the top row. The prefix in example (4) is used for a third person masculine singular proto-agent acting on a first person plural proto-patient.

(4) *shukwa-hlo·li-he?
  3M.SG->1PL-tell-HAB
  'he tells us'

'Transitive' prefixes mark two animate arguments, as in the example in (4), repeated from (2). In Table 5.1, the properties of the proto-agent that are marked by transitive prefixes are given in the leftmost column and properties of the proto-patient are given in the top row. The prefix in example (4) is used for a third person masculine singular proto-agent acting on a first person plural proto-patient.

(5) *wa-ha-yA-ka·ne?
  FACT-3M.SG.A-go.out-PNC
  'he went out'

(6) *wa-ha-yA-tho·?
  FACT-3M.SG.A-plant-PNC
  'he planted (it)'

(7) *lo-nolu:se-he?
  3M.SG.P-lazy-HAB
  'he is lazy'

<table>
<thead>
<tr>
<th>Prefix</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>*shukwa-hlo·li-he?</td>
<td>3M.SG-&gt;1PL-tell-HAB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*wa-ha-yA-ka·ne?</td>
<td>FACT-3M.SG.A-go.out-PNC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*wa-ha-yA-tho·?</td>
<td>FACT-3M.SG.A-plant-PNC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*lo-nolu:se-he?</td>
<td>3M.SG.P-lazy-HAB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The example in (6) shows that in dyadic verbs only animate arguments are marked. But since all verbs must have a pronominal prefix, verbs without any animate arguments default to an agent or patient feminine-zoic singular prefix, as shown in (8). In other words, neuter gender is a semantic category only; it is not relevant morphologically (see Koenig and Michelson 2012). In Table 5.1 the default feminine-zoic prefix is given in a cell labeled 'i-stem'.

(8) waʔ-ka-ná:nawaʔ?
FACT-3F.SG.A-get.wet-PNC
'it got wet'

All fifty-eight prefixes have varying forms (allomorphs), at least two and up to five, depending on the initial sound of the following stem. There are five stem-classes: C-stems, i-stems, o-/u-stems, e-/α-stems, and a-stems. In addition, thirty-nine of the fifty-eight prefixes have two variants depending on what occurs to their left (e.g. whether they occur word-initially or not). Allomorphy is exemplified in (9) with L39 khe-/khey- and with L33 luwa/-luwa/-luway/-luw-/huwa/-huwa/-huway­/-huw-. In total, there are 326 allomorphs, with an average of about five allomorphs per prefix. Note that the allomorphic khe- of L39 occurs with stems that begin with a consonant or i, while the allomorph khey- occurs with o/u-stems, e-/α-stems, and a-stems. This is one possible grouping, i.e. C- and i-stems, versus o/u-stems, e-/α-stems, and a-stems. The allomorphs of prefix L33 luwa/-luwa/-luway/-luw-/huwa/-huwa/-huway­/-huw- exhibit a different grouping, namely e-/α-stems and a-stems (-huw-) versus o/u-stems (-huway) versus i-stems (-huwa) versus C-stems (-luwa). There are eleven such groupings.

(9) a-stem
waʔ-khey-atnuhtuht-e?
FACT-1SG>3-wait.for-PNC
'I waited for her or them'
e-/α-stem
waʔ-khey-αhahs-e?
FACT-1SG>3-belittle-PNC
'I belittled her or them'

5.2 Evaluating paradigmatic complexity

As we mentioned in the introduction, our goal in this chapter is to delineate a form of complexity only morphology exhibits, what we call paradigmatic complexity. We will anchor our discussion of paradigmatic complexity to the number and kinds of rules needed for realizing the morphological distinctions expressed by Oneida pronominal prefixes. By discussing the issue in the context of rules for realizing morphological feature bundles (so-called rules of exponence, see Stump 2001), we can more easily provide possible measures of paradigmatic complexity and evaluate Oneida pronominal prefixes on these possible measures. We provide more speculative remarks on how and why these measures can serve as indices of paradigmatic complexity in the conclusion.

Since pronominal prefixes encode properties of participants in the event described by a verb, one can think of the proper use of pronominal prefixes by speakers as the result of the correct application of two sets of rules (or constraints). The first set of rules maps semantic properties of participants in the described situation onto morphosyntactic feature sets; the second maps morphosyntactic feature sets onto phonological marks. The first set of rules relates the semantic categories of participants in the situation types described by verbs and the values of the morphological AGR attribute; the second relates the values of the morphological AGR attribute to the phonological reflexes of those values. We have nothing to say about the first set of rules, except to note that that set of rules is simple in the case of Oneida, as the morphosyntactic features are easily inferable from observable properties of participants in situations. In other words, Oneida prefixes exhibit what one could
call semantic \( \phi \) feature sets. Pronominal prefixes reference properties of one or two animate participants in situations, number marking corresponds to model-theoretic plurality, gender marking to model-theoretic gender, and so forth.\(^4\)

So, what makes the Oneida pronominal prefix system complex? One way to approach this question is to think about what inflectional rules are. At a conceptual level, we conceive of inflectional rules as in (10a); a very informal example from Oneida is provided in (10b) (see below for more formal examples).

\[(10)\]

\[\text{a. Morphosyntactic Feature Set } \Rightarrow \text{ Form}\]

\[\text{b. 3.M.SG>1PL } \Rightarrow \text{ shukwa}\]

With (10) in mind, we can distinguish at least four kinds of complexity:

1. \text{Size:}

   \(\text{(a) What is the number of rules of the form (10)? Obviously morphological complexity increases with the number of inflectional rules. For Oneida pronominal prefixes, we need at most 326 rules (see Zwicky 1985 and later in this section for more details), since there are 326 forms (or allomorphs).}\]

   \(\text{(b) What is the size of the morphosyntactic feature space? In other words, what is the number of distinctions that can be marked (abstracting away from neutralization)? Assessing that number depends on the linguistic model of the morphosyntactic feature set, as explained later in this section.}\)

2. \text{Semantic ambiguity:} On average, how many semantic distinctions are expressed by the same form? In the case of Oneida pronominal prefixes, on average, how many combinations of participant categories are expressed by a single prefix?

3. \text{Directness:} Can we account for all the forms with rules that have the form in (10)? The form of inflectional rules in (10) is simple; it assumes a direct link between morphosyntactic feature sets and phonological forms. The question is whether a particular inflectional block (in this case, Oneida pronominal prefixes) needs anything more than that. There is evidence, as we elaborate later, which suggests that the association between bundles of morphosyntactic features and forms can be indirect and in some cases the kind of rule that is required is one where the output (form) of one rule is extended to, or identified with, the output (form) of another rule (i.e. requires reference to a function, akin to a paradigm function in Stump 2001).

4. \text{Segmentation and generalizability:} How difficult is it to recognize a pronominal prefix that is present in a particular verb form? For example, given the form \(\text{shukwahlo'lihe'}\) 'he tells us,' given earlier in (4), how easy is it to separate the prefix from the stem so that speakers can then identify the prefix in \(\text{lakhlo'lihe'}\) 'he tells me?'

\(^4\)As can be expected, this is an oversimplification, but the extent to which features do not correspond to model-theoretic properties can be attributed to ordinary grammar 'leakage.'
realistic measure of how size may lead to the complexity of the Oneida pronominal prefixes is the minimum number of morphological distinctions that may be referenced phonologically if one adopts the most parsimonious and motivated analysis of the space of feature combinations.

In Oneida, as in most, if not all, languages, there are restrictions on the combination of properties of φ feature bundles. For example, gender is a relevant morphosyntactic feature only for third person. Some of these restrictions are typologically, or logically, expected (the one we just mentioned or the fact that the person value inclusive requires dual or plural number). Some are idiosyncrasies of the morphology of Oneida, such as the fact that the feminine gender occurs only in the singular. (As mentioned earlier, when referring to two or more females, the feminine-zoic gender is used.) Formally, restrictions on the space of possible combinations of φ features result from two kinds of mechanisms or constraints: type (or sort) appropriateness conditions and feature co-occurrence restrictions. The first set of constraints encodes systematic restrictions on categories. For example, (11b) says that if the nominal index is of type 3-n-indef-index (denotes a discourse referent that is third person and not required to be unspecified/indefinite), then and only then the feature GEND is appropriate. By restricting features to the appropriate categories of nominal indices, type appropriateness conditions are a way of encoding that, as far as Oneida's morphology is concerned, gender is an attribute that only makes sense for non-indefinite third person participants. The second kind of constraints model idiosyncratic restrictions on combinations of properties. Thus, (12b) says that if a nominal index is of feminine gender, the value of the NUM attribute is singular. It so happens that dual or plural number and feminine gender are incompatible in Oneida. Feature co-occurrence restrictions encode this unpredictable incompatibility of attribute values.

\begin{itemize}
  \item (11) a. Only third person nominal indices that are not indefinite bear gender information.
  \item b. 3-n-indef-index $\Rightarrow \begin{bmatrix} \text{GEND} \text{ gender} \end{bmatrix}$
\end{itemize}

\begin{itemize}
  \item (12) a. Feminine nominal indices are singular.
  \item b. $\begin{bmatrix} \text{GEND} \text{ fem} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{NUM} \text{ sg} \end{bmatrix}$
\end{itemize}

The net effect of type appropriateness and feature co-occurrence constraints on combinations of φ features is to reduce the number of combinations from forty-nine (if features were truly orthogonal to each other) to nineteen, shown in Table 5.2.

But the number of possible φ feature combinations is further reduced by one general constraint. Participants in described situations that are inanimate are never referenced morphologically. Thus, one can speak of four semantic genders in Oneida (and Iroquoian, in general), namely, masculine, feminine, feminine-zoic, and neuter, but only three morphological genders. We dub this constraint the Animate Argument Constraint. It is stated in (13). (Recall that verbs that have only inanimate (neuter) arguments default to the feminine-zoic gender.) The upshot of this constraint is that out of the nineteen semantic categories of nominal indices, only sixteen are morphologically relevant.

\begin{itemize}
  \item (13) All and only indices for animate semantic arguments of verbs are members of the value of the AGN attribute.
\end{itemize}

All in all, linguistic analysis allows us to reduce the space of possible feature combinations from 2,499 to $16 \times 16 + 2 \times 16 = 288$ possible combinations. In addition, for first person acting on first person, and for second person or inclusive acting on second person, a reflexive construction (prefix) is used, and this further reduces the number of possible combinations from 288 to 248. Quite large, but not inconceivable.\(^5\)

\subsection*{5.4 Semantic ambiguity}

Reducing the number of rules by reducing the number of possible φ combinations (from 2,499 to 288/248, in this case) will always lead to a reduction in complexity, as it simply reduces the space of morphosyntactic properties to reference morphologically. However, a reduction in the number of possible antecedents of rules of exponence (and consequently a reduction in the number of rules) does not result in a reduction of complexity unequivocally. This is because a reduction in rule antecedents is possible only because not all of the possible feature combinations are realized by distinct forms and this fact results in semantic ambiguity, as we explain in this section.

\(^5\) Another way of measuring the complexity that may arise from the sheer size of Oneida pronominal prefixes paradigm is to use Carstairs' (1983) approach to complexity and compute the number of logically possible paradigms from the number of affixes and the number of allomorphs for each affix. The number of such 'possible paradigms' for the Latin nominal declension is a little over $3 \times 10^4$; the number of such 'possible paradigms' for Oneida pronominal prefixes is a little under $4 \times 10^5$. Whether Carstairs' measure is useful or not is a matter of debate (see Blevins 2004), but the difference in size indicates that Oneida pronominal prefixes are in another league from the Latin nominal declension.
Although the total number of morphosyntactic distinctions which can be marked in Oneida is 288/248, only fifty-eight of those are marked. To model the reduction from the number of potential morphosyntactic distinctions to the number of actual morphosyntactic distinctions, we make use of underspecification. Technically, underspecification in our rules of exponence is achieved by letting rules of exponence make reference to more or less specific types of nominal indices or properties of nominal indices. So, a rule that applies to all animate participants will have the value of the GEND attribute of the corresponding nominal index be animate, but a rule that only applies to masculine gender participants will have the value of the GEND attribute of the corresponding nominal index be masc. The hierarchies of types of nominal indices as well as the hierarchies of feature-values relevant for Oneida morphology are presented in Figures 5.1–5.4, where each non-leaf node in a hierarchy represents a more general type referred to by at least one rule of exponence.6

An example of an underspecified rule of exponence is given in (14).

(14) a. If a stem belongs to the consonant class and references a first person exclusive dual or plural proto-agent acting on a third person feminine singular proto-patient, prefix yakhi- to its phonological form.7

b. $\left[\begin{array}{c}
\text{PDGM} \\
\text{MORPH}
\end{array}\right]_{\text{AGR}} \left[\begin{array}{c}
PERS \text{excl} \\
\text{NUM} dp
\end{array}\right]_{\text{NUM} sg} \Rightarrow \text{expo (yakhi $\otimes$ 1, 2)}$

The antecedent of this rule applies to all lexemes that are in the consonant stem class and that describe situations where a first person exclusive dual or plural set of entities acts on a third person feminine singular entity. The antecedent leaves underspecified the number of the proto-agent argument by having as value for the number attribute the non-leaf type dp. In other words, the type dp covers all non-singular numbers, i.e. both plural and dual participants. The consequent simply concatenates the pronominal prefix yakhi- to the consonant-initial stem. Now, the number of distinct values of the AGR features that serve in antecedents of rules of exponence is fifty-eight, and each of the fifty-eight distinct AGR values is associated with a set of allomorphs. So, in addition to the rule in (14), we also have the rule in (15), which targets the same AGR value (that is, references the same participant categories), but applies to lexemes that belong to the a-, e-/A-, and o-/u-stem classes.

6 By Excl in Figure 5.2 we mean first person to the exclusion of second person or second person to the exclusion of first person.

7 The prefix yakhi- also applies when the proto-patient is third person indefinite/unspecified, or masculine or feminine-zoic dual or plural as the result of the application of three distinct rules of referral, as we mention later on.
82  Jean-Pierre Koenig and Karin Michelson

(15) a. If a stem belongs to the a, e/a, or o/u class and references a first person exclusive dual or plural proto-agent acting on a third person feminine singular proto-patient, prefix yakihi-

to its phonological form.

As alluded to earlier, to avoid committing ourselves to a morpheme-based view of inflection, we will call morph a class of allomorphs that share AGR values, that is, a class of consequents of rules whose antecedent AGR values are identical. For purposes of presentation, we will refer to morphs by their consonant-stem allomorphs. So, the consequents of (14b) and (15b) belong to the same morph. The set of distinct cells in Table 5.1 (or Table 6 in Lounsbury 1953) is the set of morphs in Oneida. Allomorphs of twenty-three of the fifty-eight morphs realize underspecified AGR values. Underspecification allows us to further reduce the number of morphs from the consequents of (14b) and (15b) to the same morph. The set of distinct of presentation, we will refer to morphs by their consonant-stem allomorphs.

5.5 Directness

Rules such as (14) and (15) directly associate a morphosyntactic feature set with a form. The question is whether the realization of all morphosyntactically relevant AGR values takes this direct form in Oneida. The answer is No. There are good reasons to posit rules of referral of the kind Zwicky (1985) discusses. The rules of exponence we have talked about until now have used underspecification of values of a portion of the AGR attribute to model syncretism. Implicit in this use of underspecification is the assumption that the relevant underspecified (nominal) index makes semantic sense. For example, the exponence rule that references 3M.DP acting on 3R.SG (the rule for the prefix shakoti-) can underspecify the dual/plural number distinction of the proto-agent index because non-singularity is a semantically coherent category. In contrast, there is no semantically coherent category corresponding to participants that are either third person indefinite or nonspecific (irrespective of number and gender) or third person feminine singular. The absence of a semantic natural class that would group together these two semantic indices is modelled in our account through the absence of an underspecified index that would cover these two indices.

To account for the formal similarity of exponents of third person indefinite/nonspecific and third person feminine singular arguments, we could, of course, use a disjunction in the AGR value of rules of exponence, which amounts to having two
sets of rules of exposition \((p \lor q) \rightarrow r\) is logically equivalent to \((p \rightarrow q) \land (p \rightarrow r)\), one for AGR values that include third person feminine singular indices, and one for AGR values that include third person indefinite/nonspecific. Such an analysis amounts to treating the similarity in exposition as a phonological accident. And certainly those kinds of accidental formal identity arise in inflectional systems. But, in the case of \(3\text{INDEF}\) and \(3\text{.SG}\) this is unlikely since the identity of exposition is systematic:

All prefixes that reference a \(3\text{INDEF}\) participant are formally identical to the prefixes referencing a corresponding \(3\text{.SG}\) participant (see Chafe 1977 for a possible historical explanation for that systematic formal identity).

To model what is not an accidental formal identity of exponents we use rules of referral that systematically relate the exposition of one feature bundle to that of another feature bundle. Of course, the need for rules of referral or how to model them is well known (see Stump 2001 for a detailed overview). What interests us here, though, is that rules of referral introduce another, more formal, source of complexity, something that has not been stressed before. In particular, representing rules of referral requires us to mediate the association between a morphosyntactic feature set and a form by making explicit reference to a function that takes the morphosyntactic feature set as input and outputs the form. In plain English, rather than having rules that say something like 'spell out AGR value \(a\) as \(x\)', we now have rules that say 'the value of the function that spells out \(a\) is the same as the value of the function that spells out \(β\).'

This is because a rule of referral basically says that the form marking a feature set \(α\) is the same as the form marking a feature set \(β\). In other words, it says that the values of an (exposition) function are the same. Hence, rules of referral cannot be stated without the introduction of such exposition functions. The need for this additional level of abstraction (i.e., reference to exposition functions) is an additional complexity of the Oneida pronominal prefix system.

There is evidence for five such rules of exposition in Oneida. One such rule is stated in (16). (16) says that intransitive prefixes of the A(gent) class that realize the AGR value \(< a >\) are the same as transitive prefixes that realize the AGR value \(< a >\), feminine-zoic sg \(>\), provided \(a\) is a speech participant or third masculine singular. An example prefix for this rule of referral is \(L\text{5 twa}-\), which attaches to consonant-initial stems and marks either first person inclusive plural A(gent) participants, or a first person inclusive plural proto-agent acting on a third person feminine-zoic singular proto-patient. A similar rule says that intransitive prefixes of the P(attent) class that realize the AGR value \(< a >\) are the same as transitive prefixes that realize the AGR value \(< \text{feminine-zoic sg. } a \text{ > , provided } a \text{ is not third person masculine dual/plural or third person feminine-zoic dual/plural} (\text{in which case the prefixes loti- and yoti-})\), respectively, are used).

(16) a. The pronominal prefix for stems that reference a single A(gent) speech participant or third masculine singular nominal index \(a\) is the same as the pronominal prefix for stems that reference a proto-agent \(a\) acting on a third person feminine-zoic singular proto-patient.

\[
\text{Inherent in rule (16) is the assumption that this is a case of asymmetric neutralization, as defined in Stump (2001) and Baerman et al. (2005). Evidence for the asymmetry comes from two prefixes, } L\text{7 sni- and } L\text{8 swa-} (\text{we cite, as usual, the allomorphs for consonant-initial stems}), \text{which, as transitive prefixes, neutralize the distinction between proto-agent and proto-patient arguments. When morphologically referencing two participants, sni- references either a third person feminine-} \\
\text{zoic singular proto-agent acting on a second person dual proto-patient or a second person dual proto-agent acting on a third person feminine-} \\
\text{zoic singular proto-patient. When referencing a single participant for stems that select A(gent) prefixes, sni-} \\
\text{references a second person dual participant and when referencing a single participant for stems that select P(attent) prefixes, sni-} \\
\text{likewise references a second person dual participant. The prefix swa- has a similar distribution: when referencing two} \\
\text{participants, swa- references either a third person feminine-} \\
\text{zoic singular proto-agent acting on a second person plural proto-patient or a second person plural proto-agent} \\
\text{acting on a third person feminine-} \\
\text{zoic singular proto-patient. When referencing a single participant for stems that select A(gent) prefixes, swa-} \\
\text{references a second person plural participant, and when referencing a single participant for stems that select P(attent) prefixes, swa-} \\
\text{likewise references a second person plural participant.} \text{\ (Note that sni- and swa- are the only morphs that occur in both intransitive agent and intransitive patient paradigms.) Now, if we assume the rules of exposition that correspond to sni- and swa- (and the other rules for the allomorphs of the same morph) have transitive AGR values as antecedents, the equality in (16) can derive the} \\
\text{corresponding intransitive uses of these prefixes. What is significant is the fact that the underspecified proto-role in the transitive uses of sni- and swa- is paralleled by the underspecified proto-role in three other prefixes: } L\text{31 etni-, L32 etwa-} \\
\text{and L43 yets-\text{si}. Rules of referral that extend transitive uses of sni- and swa- to intransitive uses allow us to capture this parallel and thus leads to a conceptually simpler model (albeit, not by much).} \text{\ Three additional rules of referral are needed. One 'extends' the third person feminine singular to the third person indefinite/nonspecific; the other two extend the third person feminine singular to transitive prefixes with a third person masculine}
\]
dual/plural or third person feminine-zoic dual/plural proto-agent, and to transitive prefixes with a third person masculine dual/plural or third person feminine-zoic dual/plural proto-patient.

5.6 Segmentation difficulties

The question we set out to answer in this chapter is whether there are aspects of what we call paradigmatic complexity that are quite different from the more often studied syntagmatic complexity. The paradigmatic complexity we have focused on involves a single inflectional block, and one dimension that is specific to paradigmatic complexity is the complexity that arises from combining an affix with a stem. In Oneida, there are two ways that complexity can arise from concatenating a pronominal prefix with a stem. Both are due to phonological adjustments at the boundary between pronominal prefix and stem, and both have the effect of impeding the ability of speakers to generalize from an observed verb form to all the other verb forms based on the same stem. Consider first the forms in (17), all three of which include the pronominal prefix L38 shako-, which references a third person masculine singular proto-agent acting on a '3' proto-patient. This prefix induces the deletion of a stem-initial a or i, as (17a) and (17b) indicate. As a result, the three verb forms in (17) are ambiguous as to stem class. Each of the forms could be an a-stem, an i-stem, or a C-stem. This kind of ambiguity is similar to the better-known ambiguities found in Indo-European languages, e.g. the conjugation class ambiguity of Latin amo (is it a first or third conjugation class form?).

(17) a. wa-shako-antihuht-e? ⇒ washakonuhtuhte?
   FACT-3M.SG>3-wait.for-PNC
   'he waited for her or them'

b. wa-shako-ihnúks-à? ⇒ washakohnúksa?
   FACT-3M.SG>3-fetch-PNC
   'he went after, fetched her or them'

c. wa-shako-li?wanu tú s-e? ⇒ washakoli?wanu tú se?
   FACT-3M.SG>3-ask.someone-PNC
   'he asked her or them'

Consider next the forms in (18) which show different allomorphs of the prefix L34 hak-, which indicates that the described event involves a third person masculine singular proto-agent acting on a first person singular proto-patient. The allomorph hak- occurs with e-/a-stems, while the allomorph hak- occurs with consonant stems.

(18) 3M.SG>1SG hak/hakw : C-stem or e-/a-stem?
   a. wa-hakw-Ãhahs-e?
      FACT-3M.SG>1SG-belittle-PNC
      'he belittled me'

   b. wa-hak-wahnóthahs-e?
      FACT-3M.SG>1SG-read.to-PNC
      'he read to me'

Both the types of situations exemplified in (17) and (18) lead to complexity because, in both cases, the class to which the stem belongs cannot be unambiguously determined from the forms given. As a consequence, one cannot infer in either case all the other forms in the paradigm of the stem, as selection of pronominal prefix allomorphs depends on stem class. The reason for this latter fact is specific to Iroquoian: Each morph has several allomorphs conditioned by the class to which the stem belongs, and that stem class is determined by the identity of the initial sound of the stem. In fact, the mapping between allomorphs and stem classes is itself rather complex because different morphs associate different groups of stem classes with the different allomorphs. For example, the morph referencing a first person exclusive plural agent has four allomorphs depending on the class to which the stem belongs, yakwa- if the following stem starts with a consonant, yakwa- if it starts with an i, yakw- if it starts with a, e, or A, and, finally, yaky- if it starts with o or u. The morph referencing a first person singular proto-agent acting on a '3' proto-patient, on the other hand, has only two allomorphs conditioned by the stem class, khe- before consonant-stems and i-stems and khey- before a-, e-/A-, and o/u-stems. Overall, there are eleven distinct groupings of stem classes for the fifty-eight pronominal prefix morphs. (Groupings were mentioned earlier at the end of section 5.1 in connection with the distribution of the allomorphs of prefix L39 khe- and prefix L35 luwa-.)

Why do the segmentation difficulties we have just mentioned—i.e. determining the class to which the stem belongs either because the stem-initial segment is obscured due to phonological processes or because the boundary between prefix and stem can be located in more than one place—lead to complexities? First, speakers must learn, for each morph, which group of stem classes goes with each allomorph. There are some subregularities, of course. But, ultimately, nothing can save speakers from having to learn that a-, e-/A-, and o-/u-stems require khey- as the allomorph for the 3SG>3 prefix. Second, even after having learned which allomorph of which morph goes with which grouping of stem classes, speakers cannot necessarily generalize from a given form of a verb to all the other forms of that verb.

We can quantify somewhat the complexity introduced by segmentation difficulties by asking what is the number of morphs in each stem class from which the other
fifty-seven forms of the stem class can be deduced. In a perfect system, each of the fifty-eight morphs for all five stem classes would allow one to infer the fifty-seven other forms. Table 5.3 shows that reality is far from perfect. Two notes regarding Table 5.3 are: (1) Lounsbury (1953: 56) reports that some speakers use prefixes with i-stems that are similar to those found on C-stems, while other speakers use prefixes that otherwise occur with stems with an initial vowel. The speakers that Michelson has worked with in Ontario since 1979 use the vowel-stem prefixes with only two verbs, -ihlu- 'say' and -ihuy- 'die'. The number of i-stem forms is based on the prefixes that overlap with C-stems since these represent the majority. (2) An innovation is that some of the transitive prefixes have either extended an allomorph ending in y from o-/u-stems to e/-A-stems (L33 kuway-, L48 yesay-, L49 kuway-) or developed new allomorphs in y before both o-/u-stems and e/-A-stems (e.g. L27 shikway- or L38 shakoy-). The table gives two numbers each for e-stems and A-stems; the first number is the number of morphs from which the other morphs of the same stem class can be deduced assuming the older forms without y; the second number is the number of morphs needed assuming the innovative forms with y are used instead.

Table 5.3 shows that, depending on the stem class, between 1.72 per cent and 78 per cent of verb forms allow one to infer the fifty-seven other verb forms (again, restricting ourselves for now to the combination of pronominal prefixes and stems). As Ackerman et al. (2009) have argued though, the complexity of an inflectional system partially depends on how frequent certain ambiguities are. In this particular case, the generalizability penalty is more or less severe depending on the relative frequency of the various stem classes. If C-stems, for example, are very rare, then the lack of generalizability for all but one morph is not as worrisome as if C-stems were very frequent. Table 5.4 shows the number of stems of each class in twelve pseudo-randomly chosen naturally occurring texts and the proportion of stems of each class across these twelve texts.

It seems that C-stems are the most frequent ones, accounting for 45 per cent of the stems in these twelve texts.\textsuperscript{10} So, the fact that only one of the fifty-eight forms for C-stems allow deduction of all other forms is worrisome. The average lack of generalizability from one form to all others is more severe than Table 5.3 would suggest.

5.7 Conclusion

Much of the discussion on what is a human language over the last sixty years has focused on properties of the syntactic and semantic combinatorics, the 'discrete infinity' displayed by both syntax and semantics. Measures of complexities (including, for example, where natural languages fall on the Chomsky hierarchy) have, as a consequence, focused on syntagmatic complexity, i.e. the result of putting morphemes, words, and phrases together. In this chapter, we discussed another kind of possible complexity; namely the complexity that a single position class or rule block can exhibit, what we call paradigmatic complexity. By focusing on a single block, we believe it is easier to ask questions about what makes inflection complex and explore to what extent the kind of complexity exhibited by inflection differs from syntagmatic complexities.

The first kind of complexity a single inflectional block can exhibit is the sheer range of choices one can make. In the case of Oneida pronominal prefixes this means fifty-eight morphs and 326 allomorphs (roughly five allomorphs per morph). So, not unsurprisingly, paradigmatic complexity is first a matter of number of choices. But why are inflectional choices difficult? After all, one's vocabulary is about three orders of magnitude larger than the fifty-eight morphs in the Oneida pronominal paradigm, at least according to some estimates of the average vocabulary size of English speakers. So learning or choosing words does not seem to be all that difficult. However, at least according to subjective accounts from second-language learners of English versus Oneida, learning or choosing inflected forms of a verb in Oneida is significantly more difficult than learning or choosing content words (unfortunately data suggesting that learning or using pronominal prefixes is easy or hard for native speakers is not available, as no children are learning Oneida as a first language). So, what is it that makes the retrieval and selection of a pronominal prefix difficult?

\textsuperscript{10} Despite the much more frequent occurrence of C-stems tokens in our sample texts, there is no intuitive sense in which C-stems forms are the default any more than, hypothetically, finding out that first declension forms are significantly more frequent in some Latin texts would lead us to say that first declension is the default declension class of Latin nouns.

\begin{table}[h]
\centering
\begin{tabular}{lccccc}
\hline
\textbf{C-stem} & \textbf{i-stem} & \textbf{o-/u-stem} & \textbf{e/-A-stem} & \textbf{a-stem} \\
\hline
1 & 11 & 45 & 24 & 40/20 & 30 & 26 \\
\hline
\end{tabular}
\caption{Number of morphs in each stem class from which all other fifty-seven forms can be deduced.}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{lcc}
\hline
& \textbf{Total} & \textbf{\%} \\
\hline
C-stems & 504 & 45 \\
a-stems & 373 & 34 \\
e/-A-stems & 109 & 10 \\
i-stems & 102 & 9 \\
o-/u-stems & 26 & 2 \\
\hline
\textbf{Total} & 1,114 & 100 \\
\end{tabular}
\caption{Number and percentage of stems of each class in twelve texts.}
\end{table}
We can only offer speculative remarks here. Nevertheless, we think the issue is of sufficient theoretical interest to make it worth speculating about. First, the use of a particular pronominal prefix is obligatory. In other words, given the situation being described there is a single appropriate prefix that speakers must choose. Second, in choosing or interpreting pronominal prefixes speakers and hearers must attend to several very general properties that are not necessarily the most salient participant properties in the situation (the properties are not basic level properties, so to speak).

We can quantify this putative underlying cause of perceived complexity by language learners by the number of decisions speakers must make when choosing a pronominal prefix (we leave aside allomorphy for now). They must decide if the situation involves one or two participants. If it involves one (animate) participant, they must decide if the verb selects Agent or Patient intransitive prefixes. For each participant that is morphologically referenced, they must make decisions about person and number, and for third person participants, whether it is indefinite/nonspecific or not, and, if not, what the participant's gender is.

Thus, to choose the pronominal prefix l10 la-, a speaker must have decided that the situation involved two animate participants (1), that the proto-agent was third person (2), and if not-indefinite (3) that it was masculine (4), and singular (5) and that the proto-patient was third person (6), and if not-indefinite (7) that it was feminine-zoic (8), and singular (9). In total, speakers must make, in the worst case, between four and nine decisions (each decision requires making a choice between two and four alternatives) before they can select the appropriate morph.

When comparing this number with the typical number of decisions required to select nominal or verbal suffixes in Indo-European the number is quite high. To properly decline Latin nouns speakers must only decide on the noun's number and gender (we omit case, as it is syntagmatically determined, or declension class, as it is more similar to the effect of stem-classes in Oneida). To properly conjugate, say, an ancient Greek verb only five decisions must be made (voice; mood and tense; person and number). Moreover, these five decisions are not required to select one morph, but rather a sequence of several morphs. The perceived complexity of Oneida pronominal prefixes, we surmise, is partly due to the large number of decisions speakers must make when choosing a single morph. These first two possible factors that may lead to the complexity of the Oneida pronominal prefix slot can be grouped together under the rubric attentional complexity: To make one choice (which pronominal prefix to use), speakers must attend to many distinct and concurrent properties of participants in the described situation.

Underspecification often helps speakers, as it reduces the number of decisions to be made. Thus, to choose the prefix la6 yuk- (3 > la2), speakers need only make four choices, as the prefix neutralizes all distinctions among third person except masculine singular and feminine-zoic singular. The trade-off is an increase in semantic ambiguity for hearers, as we discussed in section 5.4.

Third, because of neutralization, some properties are relevant some of the time, but not all of the time. For example, the distinction between dual and plural is never relevant for proto-patients; it is also not relevant for some proto-agents, namely just when a first or second person proto-agent acts on a '3' proto-patient. Speakers and hearers must therefore attend to the contrast between two or more first or second person proto-agents... unless the situation involves a '3' proto-patient. Furthermore, because most neutralizations are not systematic (except for the distinction between dual and plural third person proto-patients), speakers and hearers cannot rely on blanket statements on when to ignore some morphologically relevant participant properties.

Fourth, having chosen which of the fifty-eight morphs is the right one for the situation speakers must then choose an allomorph. And this is not easy. Each morph has on average five allomorphs and allomorphy, although often motivated, is not automatic; it is something that speakers must learn for each morph (and appropriately use in production). Moreover, as we have illustrated, allomorphy can lead to segmentation ambiguities for the hearer.

If number of morphs (or allomorphs) within the block leads to complexities, it would seem that any reduction in size might lead to reductions in complexities. Interestingly, we tried to show that this is not necessarily the case. We distinguished between two kinds of reduction in number of morphs. The first kind reduces the space of possible morphosyntactic distinctions to mark or reference. We showed that a skilled linguistic analysis can reduce the number of such distinctions from 2,499 for an unstructured space to 288/248 when general and particular constraints on feature and feature-value combinations are in effect. Reductions in the number of morphosyntactic distinctions that can be marked always lead to a reduction in complexity. But, the second kind of reduction is not unmitigatedly a simplification. This is the reduction from the 288/248 possible morphosyntactic combinations to the 58 Oneida pronominal morphs. We model this reduction mostly through the use of underspecification in the statement of rules of exponence. Rules of exponence can leave underspecified the type of nominal indices or features of nominal indices.

Underspecification results in a reduction in the number of rules, but at the cost of semantic ambiguity. Given a verb form with a particular pronominal prefix hearers cannot be sure of all the properties of participants in the situation that the pronominal prefix references. In certain cases, it is possible hearers do not resolve this ambiguity, but, as we showed, some of the time what is underspecified is so important to the speaker's communicative intent (what situation is being described), that hearers are most likely to have to disambiguate what is being referenced by the prefix, as when the proto-role is being underspecified.

In the end, what Oneida pronominal prefixes tell us is that there is more to linguistic complexity than the combinatorial issues we are most familiar with from work in syntax and semantics. The need for speakers to make a set of inter-related choices...
to retrieve and select a form can also lead to complexities. The importance and relevance of accessing forms when producing or understanding languages will not surprise psycholinguists. What may be more interesting is that sometimes a language’s grammatical system can make this retrieval quite a difficult matter.

Acknowledgements

We thank Karin Michelson’s Oneida collaborators, and especially Norma Kennedy. We are also grateful to Matthew Baerman, Greville Corbett, and Hanni Woodbury for reading earlier drafts of this chapter, and to Clifford Abbott, Michael K. Foster, and Hanni Woodbury for discussion of Iroquoian pronominal prefixes. We thank our colleagues Rui Chaves, Matthew Dryer, David Fertig, and Jeff Good for their input. Finally, this chapter would not have been written without Farrell Ackerman and Jim Blevins spurring our interest in this topic during their 2011 LSA Institute class.

6

Gender–number marking in Archi: Small is complex

MARINA CHUMAKINA AND GREVILLE G. CORBETT

6.1 The problem

The Nakh–Daghestanian language Archi has a small paradigm of markers realizing gender and number. Though small, this paradigm proves complex. We see this complexity in the inventory of inflectional targets, since almost all parts of speech can mark gender and number but not all lexemes within the same part of speech behave alike. Predicting which will show gender and number is not straightforward. More difficult is specifying the position of the gender and number markers: many items have infixal marking, and these are found in some instances where prefixal marking would be felicitous on purely phonological grounds. That is, Archi exhibits the typologically rare phenomenon of ‘frivolous’ infixation. We propose a number of factors which bear on the presence and position of the gender–number markers, and also on their forms and syncretism pattern; these factors overlap in ways which make it hard to isolate the impact of individual factors. Our approach will be to give the defaults, and the more specific overrides to these defaults. It is ironic that this complexity is found in the relatively small gender and number paradigm, since Archi is famed rather for the sheer size of its other paradigms.

6.2 Description of the system

For describing and analysing the morphological marking of gender–number agreement in Archi we use two main sources of data. First, there is the extensive work of Aleksandr Kibrik and his colleagues: the Archi grammar published in Russian in 1977 (here we use volumes I, II, and III, referred to as Kibrik et al. 1977, Kibrik 1977a, and Kibrik 1977b respectively) and the online collection of Archi