# Paradigmatic Generalization of Morphemes

Jochen Trommer[*]

## Abstract

In this paper, I propose a new formalism for capturing 'unnatural' syncretism (Baerman 2005): The specification of morphemes is impoverished in specific paradigm cells and hence generalized to contexts which are not subsumed by its featural content. I show that this formalism allows a systematic account of syncretism in the complex agreement systems of Ainu, Karuk (Hokan), Kulung and Limbu (Kiranti), and gives rise to a simple learning mechanism which makes substantial predictions on the subanalysis and morpheme specification for inflectional paradigms.

## 1. Introduction

It is a pervasive feature of natural language morphology that formatives which have a well-defined meaning (or function) in a specific part of an inflectional paradigm seem to take on a slightly different meaning in other paradigms or subparadigms. Take as a simple example the Ainu agreement prefix *eci-* which is consistently used to mark 2nd person plural for subjects and objects (in the following verb paradigms, the rows code features of transitive subjects, the columns features of objects; intransitive subjects are listed in alignment either with the transitive subject, or – in ergative languages – the absolutive argument (the object), and indicated by "–" for the missing coargument):

[*] For discussions leading to the present version of this paper, I am grateful to Gereon Müller, Eva Zimmermann, Daniela Henze, and Sebastian Bank. The research documented here was supported by a DFG grant to the project *Micro- and Macro-Variation: Hierarchy Effects in Kiranti and Broader Algic* (TR 521-3).

*2 in agreement*, 227–246
Sebastian Bank, Doreen Georgi & Jochen Trommer (eds.)
Linguistische Arbeits Berichte 88, Universität Leipzig 2010

(1) **Ainu Verb Agreement** (Tamura 2000)

| | **1sg** | **1pl** | **2sg** | **2pl** | **3sg** | **3pl** | – |
|---|---|---|---|---|---|---|---|
| **1sg** | | | eci- | eci- | ku- | ku- | ku- |
| **1pl** | | | eci- | eci- | ci- | ci- | -as |
| **2sg** | en- | un- | | | e- | e- | e- |
| **2pl** | ecien- | eciun- | | | eci- | eci- | eci- |
| **3sg** | en- | un- | e- | eci- | Ø- | Ø- | Ø- |
| **3pl** | en- | un- | e- | eci- | Ø- | Ø- | Ø- |

Ascribing *eci-* the feature specification [+2+pl] concisely captures its distribution in the whole paradigm except for the paradigm cells for 1st person subject and 2nd person object, where *eci-* appears to cross-reference all 2nd person objects, whether they are singular or plural. Hence in this part of the paradigm, we seem to deal with a slightly different instance (or use) of *eci-* which doesn't denote [+2+pl], but more generally [+2].

The literature knows two major approaches to "exceptional syncretism" of this type. First, the behavior of *eci-* might be captured by a rule of referral (Zwicky 1985; Stump 2001) which requires that the paradigm cells for 1sg:2sg (1st person singular subject and 2nd person singular object) are identical to the corresponding cells for 1sg:2pl forms. The problem with this approach is that it is completely stipulative, and doesn't capture the fact that the use of *eci-* in 1:2 forms although distinct from its more general use is still closely related to the "original" meaning of the prefix in the rest of the paradigm. Thus a rule of referral could with equal ease require that the completely unrelated 1:3 forms take over *eci-*. A second kind of approach which avoids this problem and which is familiar from the Distributed-Morphology literature (Noyer 1992; Halle & Marantz 1993) is maximal underspecification. For Ainu, we could assume that *eci-* is maximally underspecified and denotes [+2] throughout. This immediately captures the close connection between *eci*'s different uses, but runs potentially into another problem: *eci-* is now predicted to crossreference *all* 2nd person arguments, whether singular or plural. Since it doesn't actually appear in 2sg:3 and 3:2sg forms, it must be blocked for these cases in some way.[1] The same point can also be made from a more conceptual perspective: maximal

[1] An obvious possibility is to assume that *eci-* is blocked by the presence of the more specific 2sg marker *e-*.

underspecification combined with blocking doesn't capture the intuition that *eci-* seems to occur in a more regular and basic form for 2pl arguments compared to its extension to 1:2sg forms, an insight succinctly captured by a Rules-of-Referral analysis. Here, I propose a third, alternative approach to the behavior of *eci-* and similar cases, assuming that the formal entry for *eci-* is actually [+2+pl], but that it is generalized to [+2] in the context of transitive 1st person subjects.

## 2. The Framework

I'm assuming a paradigmatic approach to morphological exponence (cf. Wunderlich & Fabri 1994; Stump 2001) where cells of paradigms are defined as sets of fully specified feature structures such as {[Nom+1−2−pl], [Acc+2−1−pl]}. Morphemes are ordered pairs consisting of a phonological operation Φ and a (possibly underspecified) set of feature structures. Thus the Ainu prefix *ci-* which marks 1pl subjects in the context of 3rd person objects can be represented as (*Prefix*(ci), {[Nom+pl], [Acc+3]}). Since I will discuss here only concatenative (and especially affixal) morphology, I will use the simplified notation for morphemes illustrated in (2), where *ci-* is represented by (2-f). Throughout the paper I will use simple privative features for case augmented by the feature specification [+intr(ansitive)] singling out intransitive forms.

(2) **Ainu Agreement Morphemes**

| | | | | | |
|---|---|---|---|---|---|
| a. | en- | [Acc+1−pl] | e. | ku- | [Nom+1−pl][Acc+3] |
| b. | un- | [Acc+1+pl] | f. | ci- | [Nom+1+pl][Acc+3] |
| c. | eci- | [+2+pl] | g. | -as | [Nom+1+pl+intr] |
| d. | e- | [+2−pl][+3] | | | |

A morphological grammar of a language is defined by a specification of possible paradigm cells (cf. Stump 2001, a part of the formalism which I won't discuss further here), a set of morphemes such as the one in (2) for Ainu, and a list of generalization rules such as the ones in (3):

(3) **Generalization Rules for Ainu**

| | | | | | |
|---|---|---|---|---|---|
| a. | [+pl] | → | Ø | / [Nom+1][Acc+2−pl] | (eci-) |
| b. | [+3] | → | Ø | / [Nom−3−pl+intr] | (ci- & e-) |

A generalization rule basically deletes features from a morpheme - (3-a) replaces the [+2+pl] of *eci* by [+2] - in the context of a specific paradigmatic cell - thus (3-a) applies only to paradigm cells containing a feature structure subsumed by [Nom+1] and a feature structure subsumed by [Acc+2]. The realization of a paradigm cell *P* in a given language *L* is now computed as follows ((4) abstracts away from the linear order of phonological operations):

(4) **Realization of Paradigm Cells**

1. Create a Copy *C* of the morpheme set of *L*
2. Apply the generalization rules of *L* to *C* (with respect to *P*)
3. For every morpheme *M* in *C*:
   If for every feature structure *F* in *M*
   there is some feature structure in *P*
   such that *F* subsumes *P*,
   then apply the phonological operation of *M* to the stem of *P*

For example, ku-:[Nom+1–pl][+3] is inserted into the 1sg:3pl cell (specified [Nom+1–pl][Acc+3–pl]) since no generalization rule applies to the morpheme *ku-* in this context, [Nom+1–pl] in the morpheme trivially subsumes [Nom+1–pl] in the paradigm cell, and [+3] subsumes [Acc+3–pl]. For the 1sg cell ([Nom–3–pl+intr]), the morpheme ku-:[Nom+1–pl][+3] is generalized by the rule in (3-b) to [Nom+1–pl] which subsumes [Nom+1–pl+intr].

Note that the specifications of morphemes in the approach proposed here are underspecified, but only to a very limited extent. Specifications of morphemes are chosen in a way that the morpheme is not simply restricted to occur *only* in cells which are compatible with (subsumed by) it, but also occurs in *all* cells compatible with it. Thus in the terminology of Pertsova (2009), the morphemes by themselves show the most simple type of underspecification (type 1). Apparently more complex underspecification is achieved by generalization rules. Hence even though *eci-* occurs in paradigm cells containing [+2–pl] (especially 1sg:2sg and 1pl:2sg), it could not be specified as [+2] because it does not occur in *all* cells containing [+2–pl] (thus it doesn't occur in the 2sg:1pl cell). As shown by Pertsova, type1 underspecification is actually the unmarked and typologically most widespread type of underspecification in the languages of the world, an observation which is directly reflected in the structure of morphological grammars as they are proposed here. That morphemes are factually restricted to this type of underspecification follows from the learning algorithm I assume, which is spelled out in section 4. In the following, I will call the set of cells which are subsumed by a morpheme *M* in a given paradigm *P* the *range* of *M* (in *P*), and every range *R* of a morpheme *M* which has the property that all cells in *R* contain the phonology of *M* a *Pertsova field*.

## 3. More Examples for Morpheme Generalization

In this section I discuss more extensive (and more dramatic) examples of morpheme generalization, and show that the morpheme generalization framework allows a simple and general account of these cases. Here, as throughout the paper, I restrict myself to examples from verb agreement to guarantee maximal comparability of the examples.

### 3.1. Kulung

The first example I will discuss is the intransitive and transitive verb agreement paradigm of the Kiranti language Kulung (Tolsma 2006) which are shown in (5) (*p* abbreviates plural, *d* dual, *i* inclusive, and *e* exclusive, intransitive forms are aligned with the object because Kulung has an ergative system of case marking). The table in (5) does not show segmentation into separate agreement morphemes which will be given below.[2]

(5) **Kulung Verb Agreement (Past, Surface Forms)** (Tolsma 2006)

| | **1s** | **1de** | **1pe** | **1di** | **1pi** | **2s** | **2d** | **2p** | **3s** | **3d** | **3p** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1s** | | | | | | -na | -nci | -ni | -u | -uci | -uci |
| **1de** | | | | | | -Ø | -ci | -ni | -cuka | -cuka | -cuka |
| **1pe** | | | | | | -Ø | -ci | -ni | -umka | -umka | -umka |
| **1di** | | | | | | | | | -cu | -cu | -cu |
| **1pi** | | | | | | | | | -um | -um | -um |
| **2s** | -o | -cika | -ika | | | | | | -u | -ci | -ci |
| **2d** | -oci | -cika | -ika | | | | | | -cu | -cu | -cu |
| **2p** | -oni | -cika | -ika | | | | | | -num | -num | -num |
| **3s** | -o | -cika | -ika | -ci | -i | Ø | -ci | -ni | -u | -ci | -ci |
| **3d** | -o | -cika | -ika | -ci | -i | Ø | -ci | -ni | -ci | -ci | -ci |
| **3p** | -o | -cika | -ika | -ci | -i | Ø | -ci | -ni | -ci | -ci | -ci |
| **–** | -o | -cika | -ika | -ci | -i | Ø | -ci | -ni | Ø | -Ø | -Ø |

Since the morphemic composition of affix combinations in Kulung is heavyly obscured by morphophonological processes, most notably deletion of vowels under hiatus, I will base my discussion on the paradigm in (6) which abstracts away from morphophonology and provides morpheme segmentation:

[2] I discuss only the past tense paradigm here. The non-past paradigm partially uses different agreement markers.

(6) **Kulung Verb Agreement (Past, Underlying Forms)** (Tolsma 2006)

| | 1s | 1de | 1pe | 1di | 1pi | 2s | 2d | 2p | 3s | 3d | 3p |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1s** | | | | | | -na | -na-c-i | -n-i | -o-u | -o-u-c-i | -o-u-c-i |
| **1de** | | | | | | -Ø | -c-i | -n-i | -c-i-u-ka | -c-i-u-ka | -c-i-u-ka |
| **1pe** | | | | | | -Ø | -c-i | -n-i | -i-u-am-ka | -i-u-am-ka | -i-u-am-ka |
| **1di** | | | | | | | | | -c-i-u | -c-i-u | -c-i-u |
| **1pi** | | | | | | | | | -i-u-am | -i-u-am | -i-u-am |
| **2s** | -o | -c-i-ka | -i-ka | | | | | | -u | -c-i | -c-i |
| **2d** | -o-c-i | -c-i-ka | -i-ka | | | | | | -c-i-u | -c-i-u | -c-i-u |
| **2p** | -o-n-i | -c-i-ka | -i-ka | | | | | | -n-i-u-am | -n-i-u-am | -n-i-u-am |
| **3s** | -o | -c-i-ka | -i-ka | -c-i | -i | Ø | -c-i | -n-i | -u | -c-i | -c-i |
| **3d** | -o | -c-i-ka | -i-ka | -c-i | -i | Ø | -c-i | -n-i | -c-i | -c-i | -c-i |
| **3p** | -o | -c-i-ka | -i-ka | -c-i | -i | Ø | -c-i | -n-i | -c-i | -c-i | -c-i |
| **–** | -o | -c-i-ka | -i-ka | -c-i | -i | Ø | -c-i | -n-i | Ø | -Ø | -Ø |

In accordance with the fact that Kulung is an ergative language, most agreement markers follow either an absolutive or an ergative-portmanteau pattern, i.e. they show agreement with transitive subjects and transitive objects. Here I restrict my discussion mainly to the affixes which exhibit an absolutive pattern. Kulung's ergative alignment is closely mirrored in the morpheme dictionary in (7). For example *-ni* occurs with all 2pl intransitive subjects and objects establishing a Pertsova field for the features [Abs +2 –1 +pl] specified in its morpheme entry (I assume that dual is specified as [–sg –pl], singular as [+sg –pl], and plural as [–sg +pl]).

(7) **Kulung Agreement Morphemes**

| | | |
|---|---|---|
| a. | -o | [Abs +1 +sg] |
| b. | -c | [Abs –3 –sg –pl] |
| c. | -i | [Abs –3 –sg] |
| d. | -n | [Abs +2 –1 +pl] |
| e. | -am | [Erg –3 +pl][Abs +3] |
| f. | -u | [Erg–sg][Abs +3] |
| g. | -na | [Erg +sg][Abs +2] |

However, the ergative alignment of the absolutive markers is not perfect. Thus *-n* also extends to some forms which are 2pl, but for the transitive subject, not for the object. Especially, it appears as a subject marker with 1sg and all 3rd person objects. This is captured by the generalization rules in (8-a,b) which delete the Abs feature for the morphemes in (7-a-d) for the cells with the relevant object features, thus allowing insertion for all of these agreement markers for transitive subjects (note that (8-c) and (8-d) each collapse two atomic rules, deleting [–3] and [–pl] (or both) in the respective context).

(8) **Generalization Rules for Kulung**

| | | | | |
|---|---|---|---|---|
| a. | [Abs] | → | Ø | [Abs +1 +sg] |
| b. | [Abs] | → | Ø | [Abs +3] |
| c. | [–3],[–pl] | → | Ø | [Erg +3][Abs +3] |
| d. | [–3],[–pl] | → | Ø | [Erg +sg][Abs +3] |

Especially the forms with 3rd person objects nicely illustrate that generalization rules are in principle very general, i.e. not morpheme-specific. Thus all four absolutive agreement markers of Kulung are in principle blocked for transitive subject agreement (with 1st person exclusive, inclusive, and 2nd person objects), but extended to subject agreement in X:3 forms.

Morpheme generalization also provides a simple solution to the long-standing problem of *-c* which is notorious in Kiranti linguistics for exhibiting a janus face, acting as a dual marker in part of the grammar and as a general non-singular affix in other ones, creating a complex pattern of microvariation among Kiranti languages (Henze 2004). This problem also emerges in the Kulung paradigms discussed here. Thus for intransitive subjects and 1st/2nd person objects, *-c* acts as an unequivocal dual marker (restricted to non-3rd person arguments) whereas it marks dual *and* plural for 3rd person objects if the subject is singular, and for 3rd person subjects if the object is 3rd person. In the morpheme generalization framework this can be simply captured by the two generalization rules in (8-c,d) which delete the [−plural] and [−3] features for *-c* in the relevant contexts, reducing its specification to c:[−sg] which is appropriately inserted for all non-singular arguments. Again these generalization rules are not specific to *-c*, but also extend to *-i* which after application of all generalization rules has the same feature specification as *-c*.

Kulung *-c* also illustrates a further property of the proposed formalism which may be termed *morpheme-oriented resource sensitivity*: A marker is inserted only once even though it would be motivated by more than one feature structure. Thus in a 3pl:3pl form, both arguments are [−sg] and we might intuitively expect to find two occurrences of *-c* in this paradigm cell, but only one occurs. In the morpheme generalization approach, this simply follows from the realization algorithm for paradigm cells in (4) which does not distinguish between one or more specifications of a feature in the characterization of paradigm cells.[3] On the other hand, both *-i* and *-c* can be inserted realizing the same input feature in a specific paradigm cell since in contrast to Distributed-Morphology accounts of morphological spellout, the algorithm in (4) does not "delete" input features (Trommer 2003b) or induce blocking of markers sharing morphosyntactic features (Embick & Marantz 2008). In other words, the approach proposed here does not show *input-oriented resource sensitivity*.

[3]Obviously this raises the question how to capture forms where the same inflectional marker is used to realize different feature sets (see Trommer 2003a and Ortmann 1999 for discussion. I assume that empirically repeated use of a morpheme in the same paradigm cell is a highly marked state-of-affairs and is made possible by multiple listing of a morpheme in the morpheme inventory of a given language.

Note finally that *-c* occurs in slightly different positions if it appears in more or less specific forms. Thus in the 1di:3sg form, *-c* precedes *-u* whereas it follows this marker in the 1sg:3pl form. In the proposed analysis *-c* is specified as [−3 −sg −pl] in the former case , but reduced to [−singular] in the latter one. The different positions follows if the linear order of morphemes is governed by the specificity of morpheme entries for Φ-features where more specific morphemes occur closer to the stem (Noyer 1992) – unimpoverished *-c* contains 3, *-u*, 2, and completely impoverished *-c* 1 relevant feature.

### 3.2. Karuk

The Californian Hokan language Karuk (Macaulay 1992, Sappir this volume, Bank this volume), which is predominantly prefixing, has a single agreement suffix *-ap* which occurs regularly with all 2pl objects, and 2sg objects if the subject is 3rd person (the affixes of intransitive forms systematically correspond to those of transitive forms with 3sg objects):

(9) **Karuk Positive Indicative Paradigm**

| **pos** | **1sg** | **1pl** | **2sg** | **2pl** | **3sg** | **3pl** | **–** |
|---|---|---|---|---|---|---|---|
| **1sg** | – | – | ná | kiik-**ap** | ni | ni | ni |
| **1pl** | – | – | ná | kiik-**ap** | ná | ná | ná |
| **2sg** | ná | kín | – | – | ʔi | ʔi | ʔi |
| **2pl** | kaná | kín | – | – | ku | ku | ku |
| **3sg** | ná | kín | ʔi-**ap** | kiik-**ap** | ʔu | ʔu | ʔu |
| **3pl** | kaná | kín | ʔi-**ap** | kiik-**ap** | kun | kín | kun |

This can be captured in the morpheme generalization approach by positing the morpheme entry in (10) and the generalization rule in (11) (the dash before [+2] indicates that deletion of [+2] is restricted to suffixal morphemes):

(10) -ap:[Acc+2 +pl]

(11) -[+pl] → Ø /__ [Nom+3][Acc+2]

However in the negative paradigm, the paradigmatic range of *-ap* is dramatically extended. Here it also expresses plurality for 2nd and 3rd person subjects, 1st person plural subjects, and 3rd person plural objects in case the subject is non-first person (except for 2sg:3pl forms):

(12) **Karuk Negative Paradigm**

| **neg** | **1sg** | **1pl** | **2sg** | **2pl** | **3sg** | **3pl** | **–** |
|---|---|---|---|---|---|---|---|
| **1sg** | – | – | kín | kiik-**ap** | ná | ná | ná |
| **1pl** | – | – | kín | kiik-**ap** | kín | kín | kín |
| **2sg** | ná | kín-**ap** | – | – | ø | ø | Ø |
| **2pl** | kaná-**ap** | kín-**ap** | – | – | **-ap** | **-ap** | **-ap** |
| **3sg** | ná | kín-**ap** | **-ap** | kiik-**ap** | ø | **-ap** | **-ap** |
| **3pl** | kaná-**ap** | kín-**ap** | **-ap** | kiik-**ap** | **-ap** | kín-**ap** | **-ap** |

In the morpheme generalization approach all additional occurrences of *-ap* follow automatically from the generalization rules in (13) which basically state that *-ap* generalizes to the function of an all-purpose plural marker under two conditions (a) if the subject is 2nd or 3rd person, and there is a participant plural argument, and (b) if the transitive subject is 3rd person:

(13) a. -[Acc +2] → Ø /__[Nom–1][–3 +pl][+neg]
b. -[Acc +2] → Ø /__[Nom+3][Acc][+neg]

3.3. Limbu

Limbu, a Kiranti language closely related to Kulung has an agreement prefix *a*-, which occurs in a fascinating variety of contexts. *First,* it marks inclusive arguments of intransitive subjects:

(14) **Limbu Intransitive Verb Agreement**

| | **1s** | **1de** | **1pe** | **1di** | **1pi** | **2s** | **2d** | **2p** | **3s** | **3d** | **3p** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **–** | -ʔɛ | -si-ge | -i-ge | **a-** -si | **a-** | kɛ- | kɛ- -si | kɛ- -i | Ø | -si | mɛ- |

*Second,* it is used for inclusive transitive subjects and objects, and *third*, it appears in most forms with a first person exclusive object and a 2nd person subject:

(15) **Limbu Transitive Verb Agreement**

| | **1s** | **1de** | **1pe** | **1di** | **1pi** | **2s** | **2d** | **2p** | **3s** | **3d/3p** |
|---|---|---|---|---|---|---|---|---|---|---|
| **1s** | | | | | | -nɛ | -nɛ-si-ŋ | -n-i-ŋ | -u-ŋ | -u-ŋ-si-ŋ |
| **1de** | | | | | | -nɛ-si-ge | -nɛ-si-ge | -nɛ-si-ge | -si-u-ge | -si-u-ge |
| **1pe** | | | | | | -nɛ-si-ge | -nɛ-si-ge | -nɛ-si-ge | -u-m-ge | -u-m-si-m-ge |
| **1di** | | | | | | | | | **a**- -si-u | **a**- -si-u-si |
| **1pi** | | | | | | | | | **a**- -u-m | **a**- -u-m-si-m |
| **2s** | kɛ- -ʔɛ | **a**-kɛ- | **a**-kɛ- | | | | | | kɛ- -u | kɛ- -u-si |
| **2d** | **a**-kɛ- | **a**-kɛ- | **a**-kɛ- | | | | | | kɛ- -si-u | kɛ- -si-u-si |
| **2p** | **a**-kɛ- | **a**-kɛ- | **a**-kɛ- | | | | | | kɛ- -u-m | kɛ- -u-m-si-m |
| **3s** | -ʔɛ | -si-ge | -i-ge | **a**- -si | **a**- | kɛ- | kɛ- -si | kɛ- -i | -u | -u-ci |
| **3d** | mɛ- -ʔɛ | mɛ- -si-ge | mɛ- -i-ge | **a**-mɛ- -si | **a**-mɛ- | kɛ-mɛ- | kɛ-mɛ- -si | kɛ-mɛ- -i | -si-u | -si-u-si |
| **3p** | mɛ- -ʔɛ | mɛ- -si-ge | mɛ- -i-ge | **a**-mɛ- -si | **a**-mɛ- | kɛ-mɛ- | kɛ-mɛ- -si | kɛ-mɛ- -i | mɛ- -u | mɛ- -u -si |

The most general characterization which can be given for *a-* in the framework proposed here is given in (16):

(16) a-:[Abs +1][+2][−sg]

To understand this entry, note that the paradigm cell realization algorithm in (4) does not require that different feature structures in a morpheme have to match distinct feature structures of a paradigm cell. Thus for a 2sg:1pe form, [+2] of (16) matches the subject features of the paradigm cell, whereas [Abs +1] and [−sg] match the object features. On the other hand, for an intransitive 1st person inclusive form, all feature structures in (16) match a single feature structure in the paradigm cell.

(16) accounts for all occurrences of *a-* except those for transitive inclusive subjects, but the entry cannot be generally made more underspecified because this would immediately extend the coverage of the morpheme to cases where it doesn't occur. For example, removing the [Abs] feature from (16) would incorrectly predict that *a-* is also used in 1:2 forms. Again, morpheme generalization rules offer a straightforward solution to this paradox. Thus the rule in (17) captures the idea that [Abs] gets irrelevant in the context of 3rd person objects:

(17) [Abs] → Ø /__[Nom-][Abs +3]

In fact there is a fourth related use of the prefix *a-* in Limbu personal pronouns, where *a-* assumes the role of a general 1st person marker also used for exclusive forms:

(18) **Limbu Personal Pronouns** (van Driem 1987:25,28)

| | **sg** | **du** | **pl** |
|---|---|---|---|
| **1ex** | a-n-ga | **a**-n-si-ge | **a**-n-i-ge |
| **1in** | – | a-n-si | **a**-n-i |
| **2** | khɛ-n-ɛʔ | khɛ-n-si | khɛ-n-i |
| **3** | khu-n-ɛʔ | khu-n-si | |

This is captured by the generalization rule rule in (19), which reduces the morpheme entry of *a-* to bare [+1]:

(19) [Abs][+2][−sg] → Ø /__[D]

## 4. Learning

The formalism developed here lends itself to an easy learning algorithm which combines the learning of segmentation and "meaning", i.e. the assignment of morphosyntactic features to morphemes. The algorithm takes a paradigm with affix strings and morphosyntactic annotation for each affix cell as its input. It successively removes parts of the 'affix strings' (i.e. strings of prefixes or suffixes stripped from their stems) from the paradigm, adding corresponding morphemes and generalization rules to the grammar until all cells contain empty strings. The basic procedure is shown in (20):

(20) **Basic Segmentation**

1. Compute the set of Potential Morphemes *SPM* for all *free forms* in the paradigm
   and choose the Optimal Potential Morpheme *OPM* from *SPM*
2. Add *OPM* to the morpheme inventory of the grammar and
   remove all instances of *OPM* from the paradigm
3. Recompute optimal potential morphemes for the *free forms* homonymous to *OPM* from the reduced paradigm
4. Add generalization rules for all PMs resulting from 3. to the grammar, and remove all instances of these PMs from the paradigm

*Basic Segmentation* is iteratively applied to a given paradigm until all paradigm cells are empty (contain only Ø-affix strings). Steps 1 and 2 in (20) compute a morpheme entry for the potential morpheme with the "best paradigmatic coverage" in the paradigm (see below for a discussion of the evaluation metrics). Steps 3 and 4 add generalization rules for the remaining cells in the paradigm where the marker computed in the first two steps is found. The *free forms* of a paradigm are the (types of) prefix or suffix strings which occur in isolation (without other affixal material) in paradigmatic cells. For Ainu the relevant free forms are shown in (21):

(21) **Free Forms in Ainu**
en-, ecien-, un-, eciun-, ku-, ci-, e-, -as

Potential morphemes are obtained by computing all possible intersections of the morphosyntactic annotations for cells containing specific free forms *F* and cells with affix strings containing *F* and retaining only feature specifications which establish a Pertsova field (i.e. a set of paradigm cells *S* such that all paradigm cells matched by the feature specification are in *S* and contain affix strings which start or end with *F*). Thus intersecting the features of the 2sg:1sg cell (with the free form *en*- and the annotation [Nom+2−pl][Acc+1−pl]) and the 2pl:1sg cells (with the string *ecien* containing *en* and the annotation [Nom+2+pl][Acc+1−pl]) results in the potential morpheme [Nom+2][Acc+1−pl] since this combination of feature structures establishes a Pertsova Field for the cells from which the entry is computed.

In the following, I will show how iterative application of (20) derives the correct segmentation and feature assignment for the Ainu paradigm. This is an especially interesting case since instances of *eci*- in Ainu could potentially be segmented differently into independently occurring *e*- and *ci*-. (22) shows for every free form in Ainu a selection of PMs (step 1 of (20)). Note that I omit some of the PMS which subsume other PMs and have no chance of becoming optimal, and disregard -*as, ecien*-, and *eciun*- because these occur only once in the paradigm (whether viewed in isolation or as substrings of other affix strings):

(22) **Potential Morphemes in Ainu**

| Free Forms | Potential Morphemes | | Range | Cardinality |
|---|---|---|---|---|
| en- | [Nom+2] | [Acc+1−pl] | 2sg:1sg, 2pl:1sg | 2 |
| | [Nom+3] | [Acc+1−pl] | 3sg:1sg, 3pl:1sg | 2 |
| | [Nom−1−pl] | [Acc+1−pl] | 2sg:1sg, 3sg:1sg | 2 |
| | [Nom−1+pl] | [Acc+1−pl] | 2sg:1sg, 3sg:1sg | 2 |
| | ([Nom−1]) | [Acc+1−pl] | 2sg:1sg,2pl:1sg, 3sg:1sg, 3pl:1sg | 4 |
| un- | [Nom+2] | [Acc+1+pl] | 2sg:1pl, 2pl:1pl | 2 |
| | [Nom+3] | [Acc+1+pl] | 3sg:1pl, 3pl:1pl | 2 |
| | [Nom−1−pl] | [Acc+1+pl] | 2sg:1pl, 3sg:1pl | 2 |
| | [Nom−1+pl] | [Acc+1+pl] | 2sg:1pl, 3sg:1pl | 2 |
| | ([Nom−1]) | [Acc+1+pl] | 2sg:1pl,2pl:1pl, 3sg:1pl, 3pl:1pl | 4 |
| ku- | [Nom−1−pl] | [Acc+3] | 1sg:3sg,1sg:3pl | 2 |
| ci- | [Nom−1+pl] | [Acc+3] | 1pl:3sg,1pl:3pl | 2 |
| eci- | [+1−pl] | [+2] | 1sg:2sg, 1sg:2pl | 2 |
| | [+1+pl] | [+2] | 1pl:2sg, 1pl:2pl | 2 |
| | [+1] | [+2−pl] | 1pl:2sg, 1sg:2sg | 2 |
| | [+1] | [+2+pl] | 1pl:2pl, 1sg:2pl | 2 |
| | [+1] | [+2] | 1sg:2sg,1sg:2pl, 1pl:2sg, 1pl:2pl | 4 |
| | [+2 +pl] | | 2pl,2pl:1sg,2pl:1pl, 2pl:3sg,2pl:3pl 1sg:2pl,1pl:2pl, 3sg:2pl,3pl:2pl | 9 |
| | [+2 +pl] | [+1] | 2pl:1sg,2pl:1pl | 2 |
| | [+2 +pl] | [+3] | 2pl:3sg,2pl:3pl | 2 |
| | [+2 +pl] | [−2] | 2pl:1sg,2pl:1pl, 2pl:3sg,2pl:3pl | 4 |
| e- | [Nom +2 −pl] | [Acc +3] | 2sg:3sg, 2sg:3pl | 2 |
| | [Nom +3] | [Acc +2] | 3sg:2sg,3sg:2pl, 3pl:2s, 3pl:2pl | 4 |
| | | [Acc +2] | 1sg:2sg,1sg:2pl, 1pl:2s, 1pl:2pl, 3sg:2sg,3sg:2pl, 3pl:2s, 3pl:2pl | 8 |
| | [+2] | [+3] | 2sg:3sg,2pl:3sg, 2s:3pl, 2pl:3pl 3sg:2sg,3sg:2pl, 3pl:2s, 3pl:2pl | 8 |

The optimal potential morpheme from a list of PMs (step 2 of (20)) is chosen by the following criteria (ranked in this order):

(23) **Selection of the Optimal Potential Morpheme**

Choose the PM with . . .

1. . . . the maximal paradigmatic range
   (covering the maximal number of paradigmatic cells)
2. . . . the minimal recall
   (instances of the free form in the paradigm not covered by the PM)
3. . . . the minimal number of feature structures
4. . . . the biggest number of segments

   (If a number of PMs are optimal for all criteria, choose one of them randomly)

From the list in (22), eci:[+2+pl] is chosen as OPM since it has the maximal range, and the maximal range criterion outranks all other criteria. Consequently, this entry and the generalization rule in (3-a) are added to the grammar and all instances of *eci* corresponding to this morpheme are removed from the paradigm (step 2 of (20)) resulting in the paradigm in (24):

(24) **Ainu Paradigm with all Instances of *eci*:[+2+pl] Removed**

| | **1sg** | **1pl** | **2sg** | **2pl** | **3sg** | **3pl** | **–** |
|---|---|---|---|---|---|---|---|
| **1sg** | | | eci- | Ø- | ku- | ku- | ku- |
| **1pl** | | | eci- | Ø- | ci- | ci- | -as |
| **2sg** | en- | un- | | | e- | e- | e- |
| **2pl** | en- | un- | | | Ø- | Ø- | Ø- |
| **3sg** | en- | un- | e- | Ø- | Ø- | Ø- | Ø- |
| **3pl** | en- | un- | e- | Ø- | Ø- | Ø- | Ø- |

By step 3 of (20) we are now searching additional *OPM*s for the free form *eci-*. We get the entry in (25):

(25) eci-:[Nom +1][Acc +2 –pl]

Crucially, (25) is not added as a morpheme to the grammar. Instead the grammar is extended by the generalization rule in (3-a) repeated in (26) which transforms eci:[+2−pl] into a minimally less specific feature structure in the context corresponding to the morpheme specification of (25):

(26) [+pl] → Ø / [Nom+1][Acc+2−pl]

By step 4 of (20) all instances of (25) are also removed from the paradigm resulting in the paradigm in (27):

(27) **Ainu Paradigm with all Instances of *eci* Removed**

| | **1sg** | **1pl** | **2sg** | **2pl** | **3sg** | **3pl** | – |
|---|---|---|---|---|---|---|---|
| **1sg** | | | Ø- | Ø- | ku- | ku- | ku- |
| **1pl** | | | Ø- | Ø- | ci- | ci- | -as |
| **2sg** | en- | un- | | | e- | e- | e- |
| **2pl** | en- | un- | | | Ø- | Ø- | Ø- |
| **3sg** | en- | un- | e- | Ø- | Ø- | Ø- | Ø- |
| **3pl** | en- | un- | e- | Ø- | Ø- | Ø- | Ø- |

Now, we have run through one iteration of Basic Segmentation and since the paradigm is still non-empty, the procedure in (20) is repeated. Because only possible morphemes which have the maximal paradigmatic range for the free form they contain have the chance to get OPM status, I will in the following only consider this subset of possible morphemes. At this point, *en-*, *un-*, and *e-* have all the same (maximal cardinality), but *un-* and *en-* fare better for the recall criterion (there are no cells containing *un-* or *en-* which are not covered by the relative potential morphemes, but there are cells containing initial *e-* not covered by the potential morpheme for *e-*, namely all cells containing *en-*), the minimal feature structure criterion (the PM for *e-* contains 2, the PM for *en-* and *un-* contain only one feature structure), and the number-of-segments criterion. On the other hand, *en-* and *un-* fare equally well for all criteria (23), so the algorithm has to choose one of them by chance. However, the choice of OPM in this case is irrelevant because the affix strings for both PMs never overlap, and so we may either assume that *en-* gets the OPM in the 2nd iteration of *Basic Segmentation*, and *un-* in the 3rd iteration, or that the opposite order applies, without any consequences for the resulting analysis.

(28) **Ainu Potential Morphemes with Maximal Paradigmatic Range (2nd Cycle)**

| Free Form | Potential Morphemes | | Range | Cardinality |
|---|---|---|---|---|
| en- | ([Nom –1]) | [Acc +1–pl] | 2sg:1sg,2pl:1sg, 3sg:1sg,3pl:1sg | 4 |
| un- | ([Nom–1]) | [Acc+1+pl] | 2sg:1pl,2pl:1pl, 3sg:1pl,3pl:1pl | 4 |
| e- | [+3][+2–pl] | | 3sg:2sg,3pl:2sg, 2sg:3sg,2sg:3pl | 4 |
| eci- | [Nom–1–pl] | [Acc+3] | 1sg:2sg,1pl:2sg | 2 |
| ku- | [Nom–1–pl] | [Acc+3] | 1sg:3sg,1sg:3pl | 2 |
| ci- | [Nom–1+pl] | [Acc+3] | 1pl:3sg,1pl:3pl | 2 |
| -as | [Nom+1+pl +intr] | | 1pl | 1 |

Hence the fourth iteration of *Basic Segmentation* faces the paradigm in (29) which does not leave much analytic choice:

(29) **Ainu Paradigm with all Instances of *en*- and *un*- Removed**

| | **1sg** | **1pl** | **2sg** | **2pl** | **3sg** | **3pl** | – |
|---|---|---|---|---|---|---|---|
| **1sg** | | | Ø- | Ø- | ku- | ku- | ku- |
| **1pl** | | | Ø- | Ø- | ci- | ci- | -as |
| **2sg** | Ø- | Ø- | | | e- | e- | e- |
| **2pl** | Ø- | Ø- | | | Ø- | Ø- | Ø- |
| **3sg** | Ø- | Ø- | e- | Ø- | Ø- | Ø- | Ø- |
| **3pl** | Ø- | Ø- | e- | Ø- | Ø- | Ø- | Ø- |

Consequently in the following iterations, all remaining affix strings are simply assigned the PMs listed in (28), flanked by the additional generalization rule in (30), which extends *ku* and *e*- to intransitive forms:

(30) [(Acc) +3] → Ø /__[+intr –pl]

At this point, the algorithm terminates because all paradigm cells are empty.

## References

Baerman, Matthew (2005): Directionality and (Un)natural Classes in Syncretism, *Language* 80(4), 807–827.

Embick, David & Alec Marantz (2008): Architecture and Blocking, *Linguistic Inquiry,* 39(1).

Halle, Morris & Alec Marantz (1993): Distributed Morphology and the Pieces of Inflection. In: K. Hale & S. J. Keyser, eds., *The View from Building 20*. Cambridge MA: MIT Press, pp. 111–176.

Henze, Daniela (2004): Syncretism in Kiranti. Ms.

Macaulay, Monica (1992): Inverse Marking in Karuk: the Function of the Suffix -ap, *International Journal of American Linguistics* 58(2), 182–201.

Noyer, Robert R. (1992): Features, Positions and Affixes in Autonomous Morphological Structure. PhD thesis, MIT.

Ortmann, Albert (1999): Affix Repitition and Non-Redundancy in Inflectional Morphology, *Zeitschrift für Sprachwissenschaft* pp. 76–120.

Pertsova, Katya (2009): Grounding the notion of systematic syncretism in learning. Ms., University of Carolina at Chapel Hill.

Stump, Gregory T. (2001): *Inflectional Morphology*. Cambridge: Cambridge University Press.

Tamura, Suzuko (2000): *The Ainu Language*. Sanseido, Tokyo.

Tolsma, Gerard Jacobus (2006): *A grammar of Kulung*. Languages of the greater Himalayan region, Brill Academic Publishers, Leiden. pdf.

Trommer, Jochen (2003a): Distributed Optimality. PhD thesis, University of Potsdam.

Trommer, Jochen (2003b): Feature (Non-)Insertion in a Minimalist Approach to Spellout. In: *Proceedings of CLS 39*. pp. 469–480.

van Driem, George (1987): *A grammar of Limbu.*. Berlin u. a.: Mouton de Gruyter.

Wunderlich, Dieter & Ray Fabri (1994): Minimalist Morphology: An approach to inflection, *Zeitschrift fü Sprachwissenschaft* 20, 236–294.

Zwicky, Arnold M. (1985): How to describe Inflection. In: *BLS*. Vol. 11, pp. 372–386.