The unique challenge of the Archi paradigm

Author(s): Greville G. Corbett


Editors: Chundra Cathcart, Shinae Kang, and Clare S. Sandy

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via http://linguistics.berkeley.edu/bls/.

The Annual Proceedings of the Berkeley Linguistics Society is published online via eLanguage, the Linguistic Society of America's digital publishing platform.
The Unique Challenge of the Archi Paradigm

GREVILLE G. CORBETT
University of Surrey

Introduction

The verbal paradigms of the Daghestanian language Archi are justly famous for their impressive size. I argue, however, that there is a more difficult problem lodged within a small and apparently simple part of the paradigm. It concerns the expression of gender and number, in their interaction with person. I present information on the large scale of the paradigm briefly, and then outline the problem of person (§1). The need, or lack of it, for a person feature in Archi has been discussed elsewhere, so that here I can simply summarize the argument (§2). If the need for a person feature is accepted, it follows that the paradigm has an unusual shape (§3). This paradigm is genuinely difficult, as I demonstrate in the main part of the paper (§4).

1 The Issue

The main reference on Archi is the four volume grammar and texts (Kibrik 1977a, b; Kibrik, Kodzasov, Olovjannikova and Samedov 1977a, b). A more accessible summary can be found in Kibrik (1998). Archi has a truly remarkable system of inflectional morphology. Kibrik (1998:466-468) calculates that a verb in principle has 1,502,839 forms. First there are tense/aspect/mood forms; if we add the

---

1 I am grateful to the BLS organizers for the invitation to BLS, and to participants for fruitful discussion. Versions of the paper were given at “Ling Lunch” at the Université Paris-Diderot, UFR de Linguistique, 17 March 2011; the DATR Entropy Day, University of Brighton, 4 May 2011; the Max Planck Institute for Evolutionary Anthropology, Leipzig, 18 May 2011; and at the Workshop on Challenges of Complex Morphology to Morphological Theory, LSA Summer Institute, Boulder, Colorado, 27 July 2011. I thank these audiences for their suggestions, particularly Lynne Cahill and Farrell Ackerman. Special thanks are due to Marina Chumakina for sharing her expertise and insight; I am also grateful to Lisa Mack, Penny Everson and Claire Turner for help in preparing the materials. The support of the ERC (grant ERC-2008-AdG-230268 MORPHOLOGY) and of the ESRC (grant RES-062-23-0696) is gratefully acknowledged.
related gerunds, participles and masdars the total is already 12 405 (but note that some periphrastic forms are included here). If we add gender and number distinctions, and case-marked forms, the total rises to 188 463. The commentative, used for indirect speech, can be formed from all personal forms, and also from the admirative, and itself has an impressive array of forms; it is also the base for further participles. The additional forms (excluding gender and number distinctions) are 107 078. When gender/number and case distinctions are included that number rises to 1 314 376 forms. When added to 188 463 this gives 1 502 839 forms in total, as shown in (1):

(1) The number of forms derived from a single verb (Kibrik 1998:468)

<table>
<thead>
<tr>
<th>plain/</th>
<th>without gender/number forms</th>
<th>with gender/number forms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>without case forms</td>
<td>with case forms</td>
</tr>
<tr>
<td>plain</td>
<td>1 725</td>
<td>12 405</td>
</tr>
<tr>
<td>commentative</td>
<td>12 603</td>
<td>94 673</td>
</tr>
<tr>
<td>TOTALS</td>
<td>14 328</td>
<td>107 078</td>
</tr>
</tbody>
</table>

Impressive though these figures are, they are not our main concern. The system depends on a small number of basic stems, from which the large array of forms can be derived, as laid out in Chumakina (2011). These basic stems, with sound files, can be found in the Archi dictionary (Chumakina, Brown, Corbett and Quilliam 2007). Let us rather home in on the agreement markers, starting with the verb:

(2) Verbal affixes marking agreement in Archi

<table>
<thead>
<tr>
<th>GENDER</th>
<th>NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>singular</td>
</tr>
<tr>
<td>I (male human)</td>
<td>w-/&lt;w&gt;</td>
</tr>
<tr>
<td>II (female human)</td>
<td>d-/&lt;r&gt;</td>
</tr>
<tr>
<td>III (some animates, all insects, some inanimates)</td>
<td>b-/&lt;b&gt;</td>
</tr>
<tr>
<td>IV (some animates, some inanimates, abstracts)</td>
<td>Ø-/&lt;Ø&gt;</td>
</tr>
</tbody>
</table>

Archi has a complex agreement system, for which see Chumakina and Corbett
(2008). The main principle is that it has ergative-absolutive syntax, with agreement controlled by the absolutive argument. There are four genders and two numbers, with rather dramatic syncretisms, as (2) shows. The form before the slash is used in prefixal position, that after the slash is used in infixal position, with the important difference being in gender II singular. The bare stem is found for gender IV singular and for genders III and IV plural.

2 Person in Archi

Our next issue is whether (2) is sufficient, or whether verb agreement also needs to make reference to person. Archi has personal pronouns, distinguishing three persons, as well as clusivity, but it is not evident that person is a morphosyntactic feature. The issue has been discussed at length elsewhere, so here I give an outline, with references to more detailed discussion.

At first sight there is no evidence for a person feature (example (3) is from Marina Chumakina’s fieldwork, (4) and (5) are from Kibrik, Kodzasov, Olovjan-nikova and Samedov (1977b:117, 121):

(3) łonnol d-as:ar-ši d-i
woman(II)[ABS.SG] II.SG-tremble.IPfv-CVB II.SG-be.PRS
‘The woman is trembling.’

(4) zon d-irč:"in
1SG.ABS II.SG-work
‘I work (woman speaking)’

(5) un hanžugur d-aq'a ?
2SG.ABS what.way II.SG-come.PFV
How did you get here? (to a woman)

In each of these the verb, whether simple or periphrastic, agrees with the absolute argument in gender and number (glossing follows the Leipzig Glossing Rules, for which see Comrie, Haspelmath and Bickel 2004, and II indicates the second gender). Consider now the plural pronouns (examples from Aleksandr Kibrik 1972 and personal communication):

(6) teb ba-q'a
3PL I/II.PL-came
‘they (human) came’

(7) teb q'a
3PL [III/IV.PL]came
‘they (non-human) came’
In the third person plural ((6) and (7)), agreement seems again to be simply a matter of gender and number. However, the first and second persons take what appears to be the wrong form: rather than the expected human plural (comparable to (6)), they have the same form as (7). So far, we have this picture:

(10) Agreement with personal pronouns in Archi

\[
\begin{align*}
\text{zon} & \quad \text{‘I’} & \quad \rightarrow & \quad \text{gender/number agreement} \\
\text{un} & \quad \text{‘you (SG)’} & \quad \rightarrow & \quad \text{gender/number agreement} \\
\text{teb} & \quad \text{‘they’} & \quad \rightarrow & \quad \text{gender/number agreement} \\
\text{nen} / \text{nent’u} & \quad \text{‘we’ (EXCL/INCL)} & \quad \rightarrow & \quad \text{bare stem} \\
\text{žžen} & \quad \text{‘you (PL)’} & \quad \rightarrow & \quad \text{bare stem}
\end{align*}
\]

Although the singular pronouns gave no evidence for person, the picture changes when we look at examples involving conjoining and resolution of feature values (Kibrik 1977b:187):

(11) zo:n-u buwa-wu q’a
1SG.ABS-and mother(II)[SG.ABS]-and [??]come.PFV
‘I and mother came.’

The first person singular pronoun, which was apparently unproblematic on its own, causes problems when conjoined (as does the second person pronoun). Two solutions have been proposed: treating the difficulties within gender (Aleksandr Kibrik’s proposal) or recognizing a morphosyntactic feature of person in Archi. In brief, the consequences are as follows:

**Option 1** (Kibrik et al. 1977a:63-64, Kibrik 1977b:186-187):
- Archi has no person feature
- the personal pronouns zon, un, nen, and žžen form a special gender
- for resolution rules (based only on gender and number), genders must be ranked, with the gender containing the pronouns ranked higher than other genders

- Archi has a person feature
the gender resolution rules are unremarkable
person resolution is standard (except that there is only the rule that persons 1 and 2 take precedence over person 3).

Clearly both options have some merit. We will not go into the detail here (for which see Chumakina, Kibort and Corbett 2007). Corbett (2012, chapter 8) makes the additional suggestion that embeddings of the values of one morphosyntactic feature in another (as implied by the first option) should be excluded in principle. For the rest of the paper, we assume option 2, and work out its interesting consequences for the morphology of Archi.

3 Proposed Verbal Agreement Paradigm in Archi

If we accept option 2, the paradigm of the verb is unusual:

(12) Gender, number and person in the Archi verb (first attempt)

<table>
<thead>
<tr>
<th>PERSON</th>
<th>NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>singular</td>
</tr>
<tr>
<td></td>
<td>plural</td>
</tr>
<tr>
<td>1</td>
<td>gender/number</td>
</tr>
<tr>
<td></td>
<td>bare stem (person)</td>
</tr>
<tr>
<td>2</td>
<td>gender/number</td>
</tr>
<tr>
<td></td>
<td>bare stem (person)</td>
</tr>
<tr>
<td>3</td>
<td>gender/number</td>
</tr>
<tr>
<td></td>
<td>gender/number</td>
</tr>
</tbody>
</table>

There is evidence for person in the first and second persons plural. Elsewhere the specification is for gender and number. This is already surprising, and there are further unusual points, which we discuss in turn.

4 What Is Special about the Archi Agreement Paradigm?

The Archi agreement paradigm is indeed remarkable. It is not the huge paradigm indicated in (1), interesting though that is, but rather the problem outlined in (12). We draw out its interest step by step.

4.1 Looking across the Lexicon, Only Some Items Agree

When a paradigm is presented, we often assume that it applies to all possible lexical items, an expectation gained from familiar languages of Western Europe perhaps. Archi is not like that. On the one hand there are some unexpected agreement targets, including adverbs. On the other hand, only some items in each part of speech show agreement at all. (13) gives data on the number of items
which have an agreement paradigm: it is derived from Chumakina, Brown, Corbett and Quilliam (2007) and was reported in Chumakina and Corbett (2008:188); the figures for adverbs have been updated following reanalysis of some items:

(13) Numbers of items showing an agreement paradigm in Archi

<table>
<thead>
<tr>
<th></th>
<th>total</th>
<th>agreeing</th>
<th>% agreeing</th>
</tr>
</thead>
<tbody>
<tr>
<td>verbs</td>
<td>1248</td>
<td>399</td>
<td>32.0</td>
</tr>
<tr>
<td>adjectives</td>
<td>446</td>
<td>313</td>
<td>70.2</td>
</tr>
<tr>
<td>adverbs</td>
<td>392</td>
<td>21</td>
<td>5.4</td>
</tr>
<tr>
<td>postpositions</td>
<td>34</td>
<td>1</td>
<td>2.9</td>
</tr>
<tr>
<td>enclitic particles</td>
<td>4</td>
<td>1</td>
<td>(25.0)</td>
</tr>
</tbody>
</table>

The numbers are surprising. In most parts of speech, it is only a minority of items which inflect for agreement features. In part it is a matter of having a stem of the right phonological shape, but according to current knowledge we require lexical specification of the items which agree or do not agree in many instances. The part of speech where we find a majority of agreeing items is the adjective; however, adjectives have a somewhat different paradigm, and are not involved in the person problem which is our main concern.

4.2 The Pattern of Cells Where There Is Agreement Can Be Remarkable

It is not sufficient to say that a lexical item agrees or not. We may have to specify which part of its paradigm is sensitive to incoming feature requirements. This is most clearly seen in the personal pronouns, given in (14):
There are several further cases not included here. The key point is that in some cells of the paradigm there is agreement (and according to the pattern in (2) though not all the forms match the verbal forms); however, this must be lexically specified, item by item and cell by cell. We shall see an example of this agreement in (22) below. For comparison with the situation in other languages of the family see Kibrik and Kodzasov (1990:220-223).

### 4.3 Where Agreement Is Possible, Almost All Items Show the Same Pattern of Gender/Number versus Person/Number

The surprising pattern summarized in (12) is not restricted to verbs. The pattern, though not the forms, are found more generally; the issue with person arises with the different agreement targets in (13); the exception, as already mentioned, is the adjective, which has a single form throughout the plural.
4.4 The Pattern of Person-Number versus Number-Gender Is Odd

When we draw a paradigm in two dimensions, say for gender and number, we imply that we may need to make reference to either of the orthogonal features independently. If we have a third feature, this should ideally have its own dimension. It will be helpful to think in those terms here, and attempt to represent the Archi paradigm with an appropriate number of dimensions. The diagram in (15) is a first attempt:

(15) The dimensions of the Archi paradigm

![Diagram of the dimensions of the Archi paradigm]

This representation is partly right, in that it makes the point that there is a third dimension involved: the first and second persons plural are special in some way. However, the place of person is not fully clear (we return to this point in §4.6). Even so, (15) suggest that we can collapse the first and second persons, since the morphosyntax never distinguishes them. That gives us a simpler representation:
(16) The Archi paradigm simplified

(16) is a better representation, but still does not capture the place of person adequately (see §4.6 below).

4.5 The Slots with the Extra Dimension Have a *Morphomic* Pattern

When some cells in a paradigm show different behaviour from the rest, the pattern may be externally justified: the split may be justified in terms of grammatical meaning (for instance, we might find that all plural cells behave differently from all singular cells), or it may be justified in terms of phonology (for instance, all cells where the stem ends in a vowel behave differently from those where it ends in a consonant). But there are also examples where the split is purely morphology-internal, or ‘morphomic’. Consider this partial paradigm from French:

(17) French *aller* ‘go’ in the present tense

<table>
<thead>
<tr>
<th>French ‘go’</th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>vais</td>
<td>allons</td>
</tr>
<tr>
<td>2</td>
<td>vas</td>
<td>allez</td>
</tr>
<tr>
<td>3</td>
<td>va</td>
<td>vont</td>
</tr>
</tbody>
</table>
Synchronously there is no external justification for the suppletive stem in the first and second persons plural of the present tense; the distribution is morphomic. Similarly, the Archi split in the paradigm is a morphomic one. The situation of the personal pronouns given in (14) is particularly interesting. The cells which show agreement have a morphomic distribution; and within each, the distribution of cells which agree in person is also morphomic. Thus we have a morphomic pattern nested under another morphomic pattern.

4.6 The Extra Dimension Defines a Small Part (the Tail Wags the Dog)

Let us return to the different dimensions of the paradigm. We noted that our earlier representation did not position person convincingly. Though person is available only in a small part of the paradigm, it determines the shape of the whole. Consider this alternative representation (18):

(18) Person as a determining feature in the paradigm

This representation clarifies the earlier problem. Person determines which other features are realized. It is a graphic case of the tail wagging the dog.

4.7 The Different Dimensions Are Not Based on Different Stems

When we find different behaviour in different parts of a paradigm, this can often be tied to different stems. Thus in Russian the past stem is sensitive to different features than the present stem. In the Archi paradigm, however, the dramatic difference we have seen based on person is not connected to any difference in the stem. The examples that we have seen, where there is no difference in stem for the section of the paradigm where person operates, are typical.
4.8 Person Is Marked by Syncretism Going across Word Classes

The syncretism we have observed between the 1/2 persons plural form and the gender III/IV plural form is not restricted to verbs. It is found with other agreement targets too. Significantly, it is found with those adverbs that agree, and here there is an overt marker. The paradigm in (19) is set out in the original gender/number format.

(19) An agreeing adverb in Archi: *dit:aw* ‘early, soon’ (gender/number forms)

<table>
<thead>
<tr>
<th>GENDER</th>
<th>NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>singular</td>
</tr>
<tr>
<td>I</td>
<td>dit:a-w</td>
</tr>
<tr>
<td>II</td>
<td>dit:a-r-u</td>
</tr>
<tr>
<td>III</td>
<td>dit:a-b'-u</td>
</tr>
<tr>
<td>IV</td>
<td>dit:a-t'-u</td>
</tr>
</tbody>
</table>

The fact of having an overt marker should dispel any doubts about the reality of the syncretism (it is not a coincidence of uses of the bare stem, for instance). We can set out the same data including the person feature, as in (20):

(20) An agreeing adverb in Archi: *dit:aw* ‘early, soon’ (person/number)

<table>
<thead>
<tr>
<th>PERSON</th>
<th>NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>singular</td>
</tr>
<tr>
<td>1/2</td>
<td>gender/number</td>
</tr>
<tr>
<td>3</td>
<td>gender/number</td>
</tr>
</tbody>
</table>

4.9 The Syncretism Makes No Sense

Having established the syncretism, we now admit that it makes no sense. Two plural forms are available, basically human plural (genders I and II) and non-human plural (genders III and IV). Given that there is no dedicated form for person, it would seem evident which of these two forms would be chosen for agreement with first and second person pronouns in the plural. As we have seen, in Archi the other form is chosen systematically.
4.10 The Feature Person Is Non-Autonomous

In the simple instances, for each morphosyntactic feature we can point to some unique form; we justify postulating tense in English by pointing to forms like *computes* and *computed*. Similarly for each value: we justify having the value past by contrasting *computed* with other forms of the verb. There are instances, however, where values are proposed based on a combination of forms specified by values of another feature. Such instances are called ‘non-autonomous’ (Zaliznjak 1973:69-74). For example, in Classical Armenian, there is no dedicated form for the accusative (Baerman 2002). However, there is a pattern determined by number, which would lead us to postulate an accusative case. Transitive verbs govern forms which are as the nominative in the singular and as the locative in the plural. We may analyse these forms as accusative, and then the accusative case value is non-autonomous, since it has no dedicated form. However, what we are proposing for Archi is more extreme. In Armenian, and similar instances there is unique evidence for the feature case, it is only the particular value (accusative) which is non-autonomous. In Archi there is no dedicated form at all for the feature person: the feature, not a particular value, is non-autonomous.

4.11 Person Is Distinguished Only in Plural, So the Feature Is Not Realized Independently

Since morphosyntactic features are orthogonal to each other, they can be realized independently of each other. For instance, we may find an opposition of number for each case value, or indeed an opposition of person for each number value. In the Archi paradigm this is not so: we find evidence for person only in the plural. Thus person in Archi is not realized independently in this sense.

4.12 Inherent and Contextual Person and Number Values Can Conflict

We may draw a distinction between inherent and contextual features (Booij 1996, following Zwicky 1986). Inherent features are realized ‘in the right place’ (as when we find nominal number realized on nouns), while contextual features are there because of agreement or government. In the example *Victoria swims*, *Victoria* is inherently singular, while *swims* is contextually singular. Exceptionally an item may have both inherent and contextual features, and their values may conflict, as in this example from the Slavonic language, Upper Sorbian:

(21) Upper Sorbian (Faške 1981:382-383; Corbett 1987)
    moj-eho       muž-ow-a       sotr-a
    my-M.SG.GEN   husband(M.SG)-POSS-F.SG.NOM   sister(F)-SG.NOM
    ‘my husband’s sister’
The head of the phrase, *sotra* ‘sister’ is inherently feminine, and it is also inherently singular. It is contextually nominative, as determined by its syntactic position. The adjective *mužowa*, derived from the noun *muž* ‘husband’, is also feminine, singular and nominative; these are all values of contextual features: the case value derives from the syntactic position, and the gender and number values are by agreement with the head noun *sotra* ‘sister’. It is the possessive *mojeho* ‘my’ which shows the great interest of the construction. It is marked as masculine, singular and genitive. There is no expected agreement controller to account for these feature values: they do not match those of the head of the larger phrase, the noun *sotra* ‘sister’. The only other candidate agreement controller is the possessive adjective *mužowa* ‘husband’s’. We know what the feature values of *mužowa* are, namely feminine and singular – which would not, of course, account for the form *mojeho*. One solution is to suggest that possessive adjectives of this type have both inherent and contextual features of number and gender, and that their values are independent of each other (see Stump 2001:15-17). In example (21), according to this solution, *mužowa* is contextually feminine and singular (through agreement with *sotra*), and is inherently masculine and singular, as for the noun *muž* ‘husband’; it is these inherent values which *mojeho* agrees with).

The key point, then, is that features may be inherent or contextual, and that the same feature may be inherent and contextual on one and the same item; the values of the features are then independent of each other and may conflict. Having established this possibility, we return to Archi. We noted in (14) that some paradigm cells of the personal pronouns allow agreement, which is illustrated here (Kibrik 1994:349):

(22) buwa-mu b-ez dit:a<b>u
    mother(II)-SG.ERG SG.III-1SG.DAT early<SG.III>

    χ^alli a<b>u
    bread(III)[SG.ABS] made<SG.III>.PFV

‘Mother made bread for me early.’

The absolutive argument is χ^alli ‘bread’, and the verb *abu* ‘made’ agrees with it in gender and number. The adverb *dit:abu* ‘early’ is also in the gender III singular form to agree with χ^alli ‘bread’: see the paradigm in (19). Most interestingly, the first person singular pronoun in the dative case, *bez* ‘to me’, agrees: it too is gender III singular (see (14)). In Archi, the dative is also the case used with verbs of emotion and perception: thus in (23) the affected agent stands in the dative, and the object of perception takes the absolutive:
The object perceived is ź‘en ‘you (plural)’, in the absolutive case. The verb agrees with it, and has the bare stem. The adverb also agrees, and as we saw in (20) it has the infixed marker t’. Now consider the pronoun ez. From (22) we know that the first person singular pronoun, when in the dative, shows agreement. It is inherently first person singular, and its agreement is 1/2 plural (shown by the bare stem). In other words, its inherent and contextual features are in conflict. This is perhaps the most remarkable point about person in Archi. The inherent and contextual feature specification “had” to conflict in this way, yet it it seemed unimaginable, and it was therefore exciting to have the grammaticality of (23) confirmed.

5 Conclusions

What is special about the Archi agreement paradigm? The mammoth size of the verbal paradigm is of course remarkable. But the structure of the small part that involves person is perhaps of even greater interest. We noted twelve characteristics of the expression of person in the Archi paradigm. Each is of some interest. Their convergence on this small part of the paradigm makes it remarkable in the extreme.

References


The Challenge of the Archi Paradigm


Greville G. Corbett
Surrey Morphology Group
School of English and Languages
Faculty of Arts and Human Sciences
University of Surrey
Guildford, UK

g.corbett@surrey.ac.uk