

REFERENCE FORMS OF ALL CORRELATES OUR ANDEAN LANGUAGES DATABASE

RECONSTRUCTIONS, SEGMENTATIONS, ALTERNATIVE FORMS, AND POSSIBLE CORRELATES

INTRODUCTION

These notes are intended as a guide to interpretation of the *Reference Correlates* in our database, and in particular as supplementary information to our main article on our comparative study of Andean languages, namely:

Heggarty, Paul, forthcoming Feb. 2005.

Enigmas en el origen de las lenguas andinas: aplicando nuevas técnicas a las incógnitas por resolver, *Revista Andina* 40. Cuzco, Peru: Centro Bartolomé de las Casas.

Only if you have already read that article will these notes will be readily understandable, for we often make reference to concepts explained in that article (such as *Reference Correlates*).

Note: This explanation of the status of reference forms does **NOT** apply to the **CHIPAYA** data in this database. See the separate section below on Status of Chipaya Data and Reference Forms.

PROVISIONAL STATUS OF THESE DATA, CALL FOR CONTRIBUTIONS

We cannot emphasise enough that the data in our database remain in certain senses provisional, and that they are not close phonetic transcriptions (see our other .pdf on transcription policy). They data will be progressively amended and the database expanded, and on the basis of the latest data we shall produce revised quantifications of similarity between the language varieties covered, and the corresponding graphical representations of the relationships between them (see my article in *Revista Andina* 40).

Nonetheless, that article is based only on relationships in lexical semantics, so the phonetic accuracy of reconstructions and transcriptions is not generally at issue, and the potential impact on our comparative results of any inaccuracies on that level should not be overestimated. For the purposes of that article, it is the fact of whether forms are or not correlate (cognate or borrowed), not the phonetic accuracy of reconstructions, that matters. Once our phonetic transcriptions are accurately completed on the basis of our recordings, a further article will be published on calculations of similarity between the Andean language varieties in their phonetics, whose results may of course be rather different to those for their lexical semantics.

We invite contributions to any of our data from any specialists, whether comments or corrections of the existing data, or offers of data-sets for Andean language varieties not yet covered in our database.

STATUS OF CHIPAYA DATA AND REFERENCE FORMS

The data included here for Chipaya are not of the same status as those for Quechua and Aymara, indeed they are treated quite differently, so the comments here on transcriptions and reconstructions

simply do not apply to our Chipaya data. Those data are presented essentially only for reference, and for their value in demonstrating quite how different Chipaya is from Quechua and Aymara, and clearly unrelated to either.

We have not conducted any research of our own into Chipaya, nor have we studied the language in any depth, and our database simply reproduces the data collected by and taken directly from:

Cerrón-Palomino, Rodolfo & Enrique Ballón Aguirre, (in preparation).

Vocabulario chipaya-castellano, castellano-chipaya.

Fondo Editorial de la Pontificia Universidad Católica del Perú: Lima.

We are therefore in no position to make any interpretations of the data we reproduce for Chipaya, and simply reproduce them exactly as per the orthography proposed for Chipaya by those authors. Nor are we in a position to consider any reconstructions, so our Chipaya 'reference correlates' have exactly the same form as the entries for the language itself, except in some of those few cases where the Chipaya form is clearly correlate with a Quechua or Aymara form.

The version of the dictionary generously made available to us in advance of publication was an early one, still in preparation and not yet fully complete, so for a small number of our list-meanings our database as yet contains no data for Chipaya. We aim to complete these with the help of the authors.

Finally, for practical reasons of limiting the database complexity and presentation on the screen, for the moment we have only allowed a single slot (or 'field', in the database) for the Chipaya entry. So where multiple synonyms are given in Cerrón-Palomino & Ballón Aguirre's dictionary for a given list-meaning, we have allowed ourselves to enter them together (separated by commas) in that single slot. Only forms that are correlate with any of the Quechua or Aymara forms for that list-meaning are not treated in this way, because in such cases the Chipaya entry does not refer to the Chipaya reference correlate slot in any case, but to the Quechua or Aymara slot.

Entering multiple synonyms in a single correlate slot for Chipaya is possible because in this case it does not have any impact on our calculations, since they are based only on degree of *overlap* in lexical semantics. Where, as in the vast majority of cases, overlap with Quechua and Aymara is null for *all* Chipaya synonyms, the similarity rating for that list-meaning remains at 0 no matter how many Chipaya synonyms are recognised.

SEPARATE REFERENCE FORMS FOR POSSIBLE CORRELATES; PLAUSIBILITY RATINGS

In some cases it is not certain whether two forms found in different language varieties are ultimately related correlates, or simply chance lookalikes such as the forms for *red*: **puka** vs. **čupika**. (This is particularly common where one form is typically Quechua, the other typically Aymara). We adopt a very particular policy on this, rating any such cases for how plausible it seems that the two forms are ultimately connected, on a number of grounds.

This is fully explained in Heggarty (forthcoming), so we do not go into it further here, except to note its consequence for our reference correlates, which is that any cases where it is not entirely certain that two forms are actually truly correlate must be distinguished as separate correlates. The possible relationship between them is given a plausibility rating in the *Reference Correlates* form of the database, and depending on which plausibility rating the programme is set to, it will treat them as not correlate, or recognise the relationship and treats them as effectively reflexes of the same correlate.

STATUS OF REFERENCE FORMS: PROTO-FORMS?

Assessments of cognacy (if not necessarily our other type of correlates, loanwords) necessarily go back to proto-forms, so in the reference records for each of our 150 list-meanings we prefer wherever possible to give reference correlates in the assumed **PROTO-FORM** of each.

However, our study is a comparative and quantitative one, rather than specifically an attempt at reconstructions. Recall that we do not at this stage put forward our transcriptions for individual languages' forms as accurate to a high level of phonetic detail, but simply to support our comparisons between Andean language varieties in their *lexical semantics*. Similarly, our reference forms are intended principally just to help identify which words are correlate with which.

We therefore do not put our reference correlates as firm proposals for reconstructions; their status for now is purely for reference. Eventually we do hope to provide firmer proposals for reconstructions within this comparative database, and to this end, again we invite and would be most grateful for any comments and corrections to our assumed 'reconstructions'.

SOURCES FOR RECONSTRUCTIONS

Moreover, we do not by any means claim that the reference correlate forms we use are our own original reconstructions of proto-forms. On the contrary, wherever they are available we have simply followed those proposed in the few existing published sources of reconstructions.

- For **QUECHUA** these are Cerrón-Palomino (1994) and Weber *et al.* (1998), though there are some differences in their approaches, not least because Weber *et al.* limit their reconstructions to Proto-Quechua I. Weber *et al.* assume [š] in their reconstructions, rather than Cerrón-Palomino's proposal for [ʃ], and Weber *et al.* reconstruct long vowels; in both cases we follow Cerrón-Palomino.
- Cerrón-Palomino (2000: 364-66) provides reconstructions not only for Quechua but also for **AYMARA**, the only source we have used for Aymara family proto-forms.

Only for those lexemes for which those sources do not give reconstructions have we sometimes attempted our own; otherwise we simply use the extant modern forms as reference correlate forms too.

Given that it remains in dispute whether to reconstruct the aspirate and ejective series of stops in Proto-Quechua, we represent the *possible* aspirate or ejective pronunciations of the proto-forms by putting the relevant symbol in brackets, so respectively (**h**) and ('): ***rup(h)a**, ***q(')uñi**. Among the various modern

Quechua varieties that do have aspirates and ejectives, there is of course considerable variation in which lexemes have these pronunciations. So in choosing which proto-lexemes to mark as possibly aspirated or ejective, we have had to follow just one of the modern varieties that has these pronunciations, simply as a reference guide: for this we have *arbitrarily* chosen the Cuzco variety as the one we happen to have the most experience of.

In our reference correlates there is potential for uncertainty and error between retroflex or non-retroflex forms of the Proto-Quechua affricates */ç/ vs. */č/, and fricatives */š/ vs. */s/. In a number of cases, particularly where a morpheme survives only in Quechua varieties that have lost the /ç/ vs. /č/ distinction, we were unable to tell for certain to which of these sounds a particular surviving /č/ form goes back to. Likewise for the /š/ vs. /s/ distinction now lost in many varieties. In such cases, for lack of sure information to the contrary, we have generally transcribed the reference correlate with /č/ and /s/.

SEGMENTATIONS BY MORPHEME

Since cognacy, and our concept of correlateness, is an attribute not of whole words but of individual morphemes, we separate morphemes from each other as different reference correlates. Only in some cases was this not necessary if two morphemes are always found together in all varieties in the database, such as the root and suffix in Quechua **mi·k(h)u·** (*eat*).

We also mark the status and use of morphemes by using dots · to show any word-internal morpheme boundary. A dot on either side of a morpheme indicates that a further morpheme is necessarily required either before or after it. So we use a dot after any morpheme that cannot finish the word but requires a further morpheme (suffix) after it; and likewise a dot before any morpheme that cannot start a word but is only a suffix that requires a root before it.

All **ROOTS** therefore start without a dot. Noun/adjective roots end in no dot, since they require no suffix: **wasi** (*house*). Verb roots, however, do end in a dot after the final vowel, since some suffix or other is always required: **ri·** (*go*). Contrast the noun and verb roots: **para rain** (noun) vs. **para· rain** (verb).

All **SUFFIXES** start with a dot. Again, they do not end in a dot if no further suffix is necessary: all noun suffixes **·cha** (diminutive), **·kuna**, **·wan**; or grammatical verb suffixes that mark person or number **·saq**, plural **·ku**, etc.. Word-internal suffixes, usually derivational verb suffixes, that always need to be followed by a further suffix do end in a dot: **·rqu·**, **·kU·** ('reflexive/progressive'). Contrast **ri·** (*go*, a verb root), **·ri·** ('inchoative', a derivational verb suffix), and **·ri** ('question topic', a word-final clitic suffix).

In some cases it is unclear whether certain segments of words are best identified as separate morphemes or not. In such cases we do segment them in our reference correlates, because this is the only system that can allow both analyses (with different plausibility ratings), hence our analysis of the Quechua words for *smoke*, which may be simply unrelated roots **q(')uŕñi** and **quntay**, or may have rather obscure origins based on a shared root **qu·**. Likewise for the unclear relationship between the forms for *cloud*: **p(h)uyu** and **pu·ku·tay**.

This illustrates that a segmentation given in our reference correlates does not necessarily imply that we necessarily put that forward ourselves as the correct reconstruction; it is simply to allow separate analyses of the data on either of the possible reconstructions (at different plausibility settings).

ALTERNATIVE AND DERIVED FORMS

Across all the Andean language varieties in our study, a single correlate may occur in various forms, various different modern reflexes of the original form of that correlate, such as in Quechua both **kimsa** and **kinsa** for *three*, or both **turi** and **tura** for *brother* (of a woman).

Such variation is normally the result of sound changes in some varieties that derived changed forms from the original one, but in some cases might simply go back to an alternation in the proto-language. Whichever is the case, and whichever form is original and which derived, are issues that do not affect the status of forms as correlate or not, and so do not affect in any way our similarity results since they are based only on that status.

It is for information only, then, that in most cases we have sought to indicate which form is the original, where it is possible to identify that with reasonable confidence, and it is that assumed original form that we use for the form of the reference correlates.

Where it is quite simple and unproblematic to identify a single phonemic form as the ancestor of all the various modern reflexes of a correlate found across the Andean language varieties, we give only that form as the reference correlates: *e.g.* for *three* we give only **kimsa**, since it is clear that regional pronunciations such as **kinsa** are derived from **kimsa**, and not the other way around.

In other cases it is rather less certain which form is to be seen as the original, and which the derived, as for example with **tura** *vs.* **turi** for *brother* (of a woman), or metatheses such as **makwa** *vs.* **mawka** for *old* (of a man). Where we really feel we cannot confidently identify which is the original form, we include them as alternative forms in the reference correlate box, separated by the sign ~, hence for example **tura** ~ **turi**.

In some cases where it might not seem clear at first sight which form, but we do feel that there are in fact reasonable grounds to identify one of them as the original, we put that form first, separated from a sample of other forms we imagine are derived from it by the sign >. So for example the reference correlate for *blood* is given as **llawar** > **yawar**; *egg* as **ruru** > **lulu(n)**. We only do this to indicate the direction of change we feel is most likely following certain general principles of sound change, so do not put these forward with any great conviction in many individual cases, for we have not carried out a detailed case-by-case investigation that might reveal a more complex history.

The most general principle we follow is that phonetic 'weakening', *i.e.* change from a closer to a more open stricture along the chain stop → fricative → approximant, is much more frequent than the rare examples of the process in reverse ('strengthening'). Among these various instantiations of this, including of course the widespread stop → fricative changes in syllable-final position in Southern

Quechua and Southern Aymara. We also generally assume harmony between sounds as a derivation from an earlier state of disharmony.

- The alternations between /s/ and /h/ are a well-known conundrum in Quechua varieties, though in most cases we follow the default assumption of the /h/ being derived by weakening from the /s/, hence **sac̣(')a** > **haç(')a** (*tree*).
- Where a semi-consonant /y/ or /w/, i.e. consonantally an approximant, alternates with a corresponding closer stricture, such as /m/ vs. /w/, we assume that the closer stricture is the original, the more open one a derived weakened form, /m/ → /w/: **marmi** > **warmi** (*woman*).
- Where a semi-consonant /y/ or /w/ alternates with another approximant, but one with a more complex articulation, such as lateral /λ/ vs. central /y/, we assume the more complex one as original, and the semi-consonant as derived from it by weakening, /λ/ → /y/: **llawar** > **yawar** (*blood*). Compare Spanish pronunciations of <ll>.
- Alternations between the two semivowels /w/ vs. /y/, which are particularly common in Southern Quechua, are assumed to be cases of an original derivations original /w/ → /y/, since this is clear in certain cases at least, such as *day*: **p'un-chaw** > **p'un-chay**. We assume a generalisation of this change (often for dissimilation) for examples such as *brother* (of a man) **wawqi** > **wayqi**.
- For /r/ vs. /l/, we have clear evidence of historically attested changes from /r/ → /l/, particularly common in Central varieties of Quechua and Aymara (as in **rima-q** → *Lima*), and we assume the same tendency even for occasional cases of /l/ in varieties that do not show the full change: e.g. *ear* as **rinri** > **linri**; *egg* as **ruru** > **lulu(n)**.
- In some cases of alternation in Quechua between palatal vs. alveolar liquids /λ/ vs. /l/ we cannot tell which is original, and one's position can depend on whether one feels both should be reconstructed as phonemes in the proto-language, or only one (but which: /λ/ or /l/?). Aymara shows similar alternations, such as the **qaλqu** ~ **qalqu** morphemes in the numerals *seven* and *eight*. (In cases where one form is clearly derived by the other by known sound changes, we only give the known ancestor form: **allqu** for *dog* (which gave Cuzco **alqu**).
- Similar alternations occur between palatal vs. alveolar nasals /ɲ/ vs. /n/, and again unless there is a clear case of derivation by a generalised sound rule in a variety (in this case those Central varieties that have lost /ɲ/ entirely) we do not take a position on the direction of change: e.g. **ɲuqa** ~ **nuqa** (*I*).
- Where alternations occur across the classes of nasals, liquids, and /r/, except for the known /r/ → /l/ change it is usually unclear which form was the original: e.g. in the Quechua root **qarwa** ~ **qalwa** (*yellow*).
- In certain grammatical suffixes there is an alternation between the vowels /u/ and /a/, governed by morphological context (before certain other suffixes, often but not always a pattern of phonetic dissimilation with an /u/ sound in that suffix). In such cases we transcribe this by the usual convention of an upper case <U>: **·kU**, **·yKU**.
- For alternations between vowels in roots we assume a general tendency for disharmony to lead to vowel harmony, so the form in which the vowels are different is assumed original: e.g. **pani** > **pana** (*sister*, of a man). In other cases there is harmony in neither form, and we do not take a position on which is the original: **tura** ~ **turi** (*brother*, of a woman).

- Likewise for multiple ejective or aspirated consonants in Aymara languages, we assume that the later ones in the root arose by harmony with the first one, as for example in *bread*: **t'anta** > **t'ant'a**.
- Cases of metathesis are legion in the Andean languages, and in few cases are there convincing grounds for identifying which form is the original, hence: **makwa** ~ **mawka** (*old*, of a man); **warma** ~ **wamra** *young*.

REFERENCES

Cerrón-Palomino, Rodolfo, 2000. *Lingüística aimara*.

Cuzco, Peru: Centro Bartolomé de las Casas.

Cerrón-Palomino, Rodolfo, 1994. *Quechua sureño: diccionario unificado*.

Lima, Peru: Biblioteca Nacional del Perú.

Heggarty, Paul, forthcoming Feb. 2005.

Enigmas en el origen de las lenguas andinas: aplicando nuevas técnicas a las incógnitas por resolver, *Revista Andina* 40. Cuzco, Peru: Centro Bartolomé de las Casas.

Cerrón-Palomino, Rodolfo & Enrique Ballón Aguirre, in preparation.

Vocabulario chipaya-castellano, castellano-chipaya.

Lima, Peru: Fondo Editorial de la Pontificia Universidad Católica del Perú.

Weber, David John & et al., 1998. *Rimaycuna – Quechua de Huánuco*.

Lima, Peru: Instituto Lingüístico de Verano (SIL).