

HOW TO USE THIS DATABASE

INTRODUCTION

This comparative database of the lexical semantics of Andean languages is based on a list of 150 basic meanings. For each of these meanings, the database includes data on the word-forms used in that meaning in 21 different varieties of Andean languages. This gives a total of 150×21 separate 'records' in the main table of the database, *i.e.* 3150 records. These are stored in the main *Comparative Data* table.

All the data themselves are stored in the '**TABLES**' of the database, as the most efficient framework in which to store them. However, the basic tables are not very suitable for visualising the data; for that it is much better to look at them in one of the three '**FORMS**' corresponding to the tables.

Note that with 3150 quite complex records, on computers with small memory and slow processors, the database will not necessarily function very quickly. The most important consequence is that some data – such as the language names and the phonemic forms of particular correlates – may appear 'blank' for a second or two while they are retrieved and processed by the programme, especially when you move from one record to the next by scrolling. If a field is still blank and you want to see it immediately, click on it.

THE *CORRELATE REFERENCE* TABLE AND FORM

As well as the main *Comparative Data* table and form, there is also another table, the *Correlate Reference* table, again with its corresponding form which we recommend you use to view these data more easily.

This the *Correlate Reference* table (and form) contains just 150 records, one for each list-meaning. In each of these 150 records are stored the 'reference' data, the collection of all the 'correlate' states (cognate or borrowed) found in any of the 21 varieties for that list-meaning. So for example for the list-meaning *FOUR*, there is one record in the Correlate Reference table, which includes the following unrelated '**REFERENCE CORRELATES**' for that list-meaning: **tawa** (Quechua), **çusku** (Quechua), **puši** (Aymara), **paqpik** (Chipaya), and **cuatro** (a Spanish loanword used in some Andean language varieties). Together these reference correlates cover all the words used to mean four in all 21 language varieties in the database. For more details, see the data samples on www.quechua.org.uk/Eng/Cpv/DataSamples.htm. For detailed information on the status of the 'transcriptions' of the reference correlates, see our other .pdf file on that subject.

The *Correlate Reference* table is also where, for each list-meaning, the details are stored on all of the sub-senses and various more indirectly related meanings that are recognised in our study. In the record for the list-meaning *SUN*, for example, are stored the data on the sub-senses recognised in our study, *celestial object* and *light/heat of the sun*; and on more indirectly related meanings such as the verb *burn*, which we recognise and include in our study because in many Andean languages this meaning is expressed by the same root that in other varieties has the meaning *SUN*.

THE MAIN *COMPARATIVE DATA* TABLE

In the main *Comparative Data* table, each of the 3150 records is composed of the same 'fields', such as:

- A reference number for which list meaning that record refers to, among the 150 total list meanings.
- A reference number for which language variety that record refers to, among the 21 total language varieties.
- A number of different fields showing which of the references correlates are found in this language variety, what particular phonemic/phonetic form they take, and all other data relevant to this list-meaning in this particular language variety.

There are two different ways of visualising these data, in either of the two corresponding **FORMS**.

VIEWING THE MAIN DATA IN THE *COMPARATIVE DATA: OVERVIEW* FORM

The *Comparative Data: Overview* form allows you to compare, listed one on top of the other on the same page of the screen, several of the 3150 records at a time, for example all the different words used for a given list meaning in multiple language varieties. So for the list-meaning *FOUR*, you can see on the same screen which Andean language varieties use the word **tawa**, which use **çusku**, which use **puši**, and so on.

Alternatively, if you select only the records for one language variety (see below for how to do this), you can see on the same screen the words used in that language variety for various different list meanings at a time, say for example all the list-meaning *ONE* to *TEN*.

VIEWING THE MAIN DATA IN THE *COMPARATIVE DATA: IN DETAIL* FORM

In order to show multiple records at once, however, the Overview form does not have space to show in detail all the data stored in a single one of the 3150 records, *i.e.* one list-meaning in one language variety. To see the data in this detail, view the main data in the *Comparative Data: in Detail* form instead.

LANGUAGE REFERENCE SYSTEM

Each of the 21 language varieties covered in our database is given a special reference number, such as for example the Quechua variety of Inkawasi (Ferreñafe, in northern Peru) has the reference number 121212001. This may seem rather complex, but it is in fact a simple system that makes it much easier to select particular interesting subsets of language varieties to view, hiding others that may not be of interest for particular purposes, and thus also speeding up the programme.

The numbering system is as follows. Each of the nine separate digits in the language reference number refers to one level in the traditional classification of Andean languages. So the first digit identifies the language family: 1----- for all Quechua varieties (hence the reference number for the Inkawasi

Quechua variety starts with the digit 1); 2----- for all Aymara (also known as Aru, or Jaqi) varieties; 3----- for Chipaya.

The second digit refers to the first major split in the family identified in each language family according to the traditional classification. So the reference numbers for all varieties of Quechua I (Central Quechua) start with 11-----, *i.e.* 1 (for Quechua), then 1 (for Quechua I); all varieties of Quechua II (North/South Quechua) start instead with 12-----, *i.e.* 1 (for Quechua), but then 2 (for Quechua I). Likewise all Central Aymara varieties start with 21-----, but all Southern or Altiplano Aymara varieties start with 22-----.

This continues through the classification, so at the next level down, Quechua IIa varieties start with 121-----, QIIb with 122-----, and QIIc with 123-----. Further details will be given in a later version of these notes, though most users are unlikely to wish to go beyond the third level in the classification.

FILTERING SUBSETS OF RECORDS BY LANGUAGES

As we noted above, the language reference numbers are particularly useful for filtering the 3150 records down to a much smaller and more manageable subset for particular purposes, which also makes the programme run faster.

Filtering is very easy. If you want to see only the Aymara language varieties, for example, find the individual digit boxes that make up the language reference number, place your cursor over the first figure in the language code, the red **Family** figure, and there right-click on the mouse. A menu will pop up with a series of options. In the [Filter For] box, type 2 (the family code for Aymara) and hit Enter. Instead of 3150 records, you will now have in your form only 750 records, *i.e.* data for the 150 list-meanings only in the five Aymara language varieties covered in our study.

If within these you only want to look at Central Aymara varieties, to compare only Jaqaru and Kawki, then you repeat this filtering process for the second figure in the language code, the **Main Branch** figure: right-click on it, and do [Filter For] the value 1 (Central Aymara). You will now be left with just 300 records in the form, for the 150 list-meanings for Jaqaru, and for Kawki.

To cancel at any stage and go back to all 3150 records, when you right-click just select 'Remove Filter/Sort'. There are plenty of more refined filtering mechanisms within Access that you can try out.

To see all 150 records for just a **SINGLE LANGUAGE** at a time, the quickest way is to right-click not on the individual digit boxes that make up the language reference number, but on the one box that gives the 'Full Language Index Number', and from the menu choose [Filter By Selection]. Conversely, to **EXCLUDE** individual languages, choose [Filter Excluding Selection].

FILTERING AND SORTING SUBSETS OF RECORDS BY LIST-MEANING

You may also wish to filter by particular list meanings, which again you do by right-clicking on the **green** list-meaning index number box. The meanings are normally listed by default in the 'elicitation

order' used in our fieldwork, which grouped words together by semantic field and by part of speech. So numerals come first, then grammatical words and pronouns, then parts of the body, and so on. All adjectives come together with index numbers from 700 to 799, and all verbs from 800 to 999.

You can, however, resort these as you wish. If you wish to sort by alphabetical order of the list meaning, in English, then right-click on the list meaning box (not its index number box), and choose either [Sort Ascending] or [Sort Descending]. To sort by alphabetical order of the list meaning in Spanish, right click on that box instead. You can filter and sort like this for most of the different fields in the tables.