

Reflections on the Huallaga Quechua dictionary: derived forms as subentries

David J. Weber
SIL International*

September 9, 2002

In 1998, together with Félix Cayco Zambrano, Teodoro Cayco Villar and Marlene Ballena Dávila, I published a dictionary of Huallaga (Huánuco) Quechua [9]. This is a Quechua dictionary *IN QUECHUA*, with definitions in Quechua, usable by a monolingual Quechua speaker. It also has equivalents, translations and indices in Spanish and English so is usable by speakers of those languages.¹

The main point of this note (as elaborated in section 2) is that a derived form normally derives from a particular sense of its base and that it may be desirable to represent this relationship in the dictionary of a language in which—like Quechua—a major proportion of the lexicon is derived.

1 Some practical issues

A dictionary² is a tool for the delivery of lexical information. No lexical database and no dictionary is complete or perfect; each represents a series of compromises. How a lexical database is stored and how it is manipulated—entered, sorted, displayed, printed—would ideally be independent, but we don't live in an ideal world.

A widely-used compromise between competing demands has been to store the lexical database as a sequence of records, each record consisting of a sequence of fields, each field consisting of a field identifier (attribute) followed by the information (value).³ Various sort programs can reorder records but within an entry information is generally printed in the order in which it is stored.⁴

*Presented to the 2002 EMELD Conference, August 3–5, 2002, Ypsilanti, MI; see <http://saussure.linguistlist.org/cfdocs/emeld/workshop2002/mainconf02.html>. I thank Gary Simons and Mike Maxwell for comments and corrections on earlier versions. But don't blame them for my indiscretions.

¹It can be purchased in Peru at Javier Prado 200; Magdalena, Lima. Sales to outside of Peru are handled by Seta Soledad Esteban of the "Liberia Interregna" bookstore, Fax 426-2742.

²I use "lexicon" (and "lexicons", not "lexica") to refer to speakers' knowledge of the lexical resources of their language, "lexical database" to refer to an information structure that represents lexical information (reflecting characteristics of a lexicon), and "dictionary" to refer to a rendering of a lexical database, whether printed on pages or displayed in some electronic form.

³Tools like Shoebox [1] and the Making Dictionaries package [4] were developed to serve this approach.

⁴This is not true for database programs that print information according to user-defined templates. However, such programs present their own challenges, most notably their proprietary data formats.

The Huallaga Quechua lexical database was stored as such a record-structured list of attribute-value pairs, but with an innovation to allow derived forms to be embedded within the entry of its base (as described in section 2). It was maintained using an emacs editor and, for two or three stages, Shoebox.

The printed dictionary was formatted with LaTeX after first applying a series of conversion programs.⁵ Beyond basic formatting, this process automated (or made as automatic as feasible) a number of features: (1) the numbering of multiple senses, (2) the formatting of subentries as described below, (3) the generation of cross-reference entries for alternate spellings and subentries (with guidance for finding them), (4) the generation of the Spanish and English indices from information given in the lexical entries,⁶ and (5) hyphenation for four different forms: Quechua as written (practical orthography), Quechua in a pseudo-phonemic representation, Spanish, and English.

And using LaTeX to format the dictionary proper (the entries) made it possible to author the other parts of the dictionary—from the front matter to the appendices—with LaTeX.

The Huallaga Quechua dictionary is good because it had pervasive, profound, and direct input from native speakers of the language, particularly that of my co-authors Teodoro and Felix. They learned some basic aspects of lexicography and dictionary making. And they learned to type, to use a computer, and to enter and edit data with various computer programs. This was possible because the database format and computer programs used were not intractable.

In some cases we intentionally kept things simpler than the ideal. Let me give a couple of examples:

1. It is normally good practice to use a distinct field identifier (attribute) for each type of information. For example, distinct field identifiers should be used for different kinds of notes: subdialectal variation, usage, ethnographic or encyclopedic information, and so forth. However, managing the different field identifiers proved to be difficult and discouraged my Quechua colleagues from adding valuable information. (I particularly encouraged them to include descriptions of Quechua beliefs and customs.) Consequently we decided to use a single field identifier for all notes. The type of note can be seen from the content; each note is relevant either to what precedes or on its own.
2. We skirted fine distinctions in semantic similarity or dissimilarity by using a simple, 3-point scale: (1) synonym or near synonym, (2) related (belonging to the same semantic subdomain but neither a synonym nor an antonym) and (3) antonym or near antonym.

⁵The lexical database was converted to Latex format by a series of programs (managed with makefile): (1) put each field on a single line (`awk`); (2) jettison fields not to be printed and put each index entry on a single line (`sed`); (3) make cross-reference entries (`awk`); (4) sort (`srt`); (5) hyphenate (i.e., introduce discretionary hyphens into) Quechua written in phonemic form (`syl`); (6) hyphenate Quechua written in Hispanic orthography (`syl`); (7) hyphenate Spanish (`syl`); (English is hyphenated by LaTeX); (8) change braces to appropriate font-changing instructions (`sed`); (9) convert the index fields to LaTeX format and number multiple senses (`awk`); (10) introduce LaTeX formatting (`sed`); (11) prepare the indices for Spanish and English by hyphenating Quechua and Spanish and introducing the correct format; (12) format the entire document (`latex`); (13) produce a Postscript file (`dvips`). (This resembles Bill Poser's process for generating a printed Carrier dictionary from a database as described in conjunction with [7]).

⁶The *Making Dictionaries* approach supports producing a "finder list" but our approach produced two-tiered indices with page numbers. These were generated entirely from information given in the lexical entries, with no human intervention (as required in the *Making Dictionaries* approach).

2 Derived forms and multiple senses

Quechua is an agglutinative language. Its genius lies in its productive word formation.⁷ The number of roots is relatively small—perhaps only a few thousand native Quechua roots and less than a thousand roots borrowed from Spanish—but from these are derived, by the addition of suffixes and by compounding, thousands upon thousands of lexemes.⁸

For example, from the base *allcha*-⁹ is derived *allchacä*-, and from this, *allchacächi*-:

allcha- *v.tr.* ‘to fix’

allchacä- *v.i.* ‘to get well, to recover from a physical or mental disorder’

allchacächi- *v.tr.* ‘to heal’

Likewise, from *warmi* ‘woman, wife’ and *ashi*- *v.tr.* ‘to seek’ is derived the compound *warmi ashi*-, and this can be nominalized with -y:

ashi- *v.tr.* ‘to search for, to seek’

warmi ashi- *v.tr.* ‘to ask a father for his daughter’s hand in marriage’

warmi ashiy *s.n.* ‘the procedure by which a family arranges the marriage of a son’

As illustrated, in the dictionary each derived form is embedded (and indented) under the form from which it is derived (its base): lexemes derived by adding a suffix are embedded under the root or stem to which the suffix is attached; those derived by compounding are embedded under the second root or stem, the head, Quechua being a head-final language.

Words, of course, have multiple senses. **A derived form normally derives from a particular sense of its base.** In large measure this is simply due to the compositional nature of newly derived forms: the meaning of the whole is a function of the meanings of its parts. There are, of course, exceptions because with time the meanings of derived forms may wander from their compositional origins.¹⁰

⁷That is not to say that the meanings of derived forms are entirely compositional; derived forms have—to one degree or another—a life of their own.

⁸Some dictionaries published for Quechua merit the criticism of England [5, pág. 34], that they represent the lexicon of a language as incomplete, thus diminishing the status of the language. England says:

“How many of us have been dismayed on hearing someone assure us that language X (in my case it was Quechua) is a primitive language, since we try so hard to dispel the notion of ‘primitive’ languages? I was much more dismayed to discover that, in the Quechua instance, the person had a seemingly legitimate reason for this idea: that there are only 5,000 words in some dictionary of the language. And who is responsible for writing the dictionary?”

This is because of a preoccupation with ROOTS at the expense of derived forms. The reasons for this are discussed in Weber [8].

⁹Even the root *allcha*- reflects prior derivation; it is the fusion of /aʎi/ ‘good’ and /-ča:/ ‘to cause to become’; Weber et al. [8, 637].

¹⁰Let me illustrate with an incident I had during my first weeks in the town of Santiago de Llacón. One morning a woman arrived to our house, with her baby, and began to plead, “Tayta lindu mishti, wamrāta ushachipāmay!” She said this again and again, each time with greater urgency.

Before having gone to Llacón I had learned some Quechua roots and dozens of suffixes, and I understood something about the morphology. Thus, I understood the root *usha*- ‘finish’ and the suffixes -*chi* ‘causative’, -*pā* ‘benefactive’, -*ma* ‘first person object’, and -*y* ‘imperative’. Accordingly, the combination should mean “Dear sir, please terminate my baby for me.” I imagined the worst; did she really want me to finish off her child?

Finally, with a streak of creativity born of desperation, and still not understanding what the woman meant,

English dictionaries do not represent the relationship of a derived form to the corresponding sense of its base. For example, the entry for *run* in *The Oxford American Dictionary of Current English* [2, 703] gives various senses of the verb *run* and then various senses of the noun *run* (derived—I assume—by “null” derivation: [_N [_V run]–∅]). Among these occur the following (with their respective numbers):

- v. ... 27 *intr.* (of hosiery) unravel along a line from the point of a snag.
- n. ... 14 a line of unraveled stitches, esp. from the point of a snag (in hosiery).

These are listed with no indication that the 14th nominal sense derives from the 27th verbal sense. And the division by part of speech obscures that some verbal senses spawn derivatives while others do not.

Perhaps for an English dictionary failing to represent this information is inconsequential, since only a modest proportion of the English lexicon is derived. But **for some languages it may be desirable to represent the relationship of a derived form to the sense of its base.** This, I suggest, is likely to be the case for languages in which—like Quechua—a major proportion of the lexicon is derived.¹¹

To represent the relationship of a derived form to the sense of its base we decided to embed subentries under the corresponding sense of the base.¹² To make this more visually apparent each level of embedding is set in a bit:

```
entry
sense
  subentry
  sense of subentry
    subsubentry
    sense of subsubentry
```

For example:

mishqui s. ‘candy; sweet, savory’

mishquichi- v.tr. ‘to deceive with the hope of getting something good’

mishquichipä- v.tr. ‘to flatter; to entice (particularly when selling something), to deceive (saying that something is good when it is not)’

I responded, “It is not permitted to a foreigner to do that.” She accepted that and left. Later I learned that *usha-chi-* means ‘baptize’. The priest had come to Llacón for his yearly visit and she was asking me to have her child baptized, to become his godfather.

It is said that *usha-chi-* (finish-causative-) originated from an old belief: Children are born as devils—horns, tail and all—so it is necessary to finish off the devilish part for the child to be human. Or something like that... it’s hard to tell for now *ushachi-* simply means baptize.

¹¹And it becomes even more apparent when considering Herrero and Sánchez de Lozada’s monumental dictionary of Cochabamba Quechua [6]. One need only scan it quickly to see that the majority of entries are derived.

¹²“Embed” suggests recursion but the limitations of some of the programs used made simple recursion impractical. Embedding was implemented by using distinct field identifiers for each level of embedding. For example, for the part of speech field, in the main entry the identifier is \ps, in subentries it is _ps (with an underscore between backslash and p), and in subsubentries it is __ps (with two underscores between backslash and p).

Storing the dictionary in this way made it straightforward to format subentries as embedded. (If they had been stored as separate records, a significantly more sophisticated process would have been required.) It is also trivial to convert the embedded entries so they print as separate entries.

(Another sense of *mishquichipä-* is ‘to eat slowly, appreciating the flavor’. From this can be derived the manner adverb *mishquichipayllapa* ‘savoring (as one eats)’.)

As expected, a particular sense (of multiple ones) may have a derived form. In the following, for example, the derived form *anaj* is based on the first sense of *ana-*, that of physical hardness, not on the second, that of difficulty. And as illustrated by the two senses of *anajyä-*, derived forms may have multiple senses.

ana-

1. *v.i.* ‘to be hard (of things that can be soft, like corn, but not with hard things like rocks)’

anaj s.a. ‘hard’

anajyä-

1. *v.i.* ‘become hard’

anajyächi- v.tr. ‘harden’

2. *v.i.* ‘to become stingy; to become hard hearted’

2. *v.tr.-na*¹³ ‘do with difficulty’

Further examples follow:

cuti- v.i. ‘to return’

cutichi- v.tr. ‘return (something to someone)’

ayñita cutichi-

1. *v.tr.* ‘help another in exchange for help given’

2. *v.tr.* ‘to take revenge, to get even with’

cutipa

1. *s.n.* ‘relapse’

2. *s.n.* ‘second cultivation of corn (done when it is over 50 centimeters high)’

cutipa- v.i. ‘to repeat (especially referring to the second cultivation of corn)’

cutipä- v.tr. ‘to return to a woman after having left her’

cutiricU- v.i. ‘to go back on one’s word, to retract an offer’

yacha-

1. *v.tr.-y*¹⁴ ‘to know how to’

yachacU- v.tr.-y ‘to learn’

yachachi- v.tr.-y ‘to teach’

yachachimu- v.tr. ‘give dictation’

yachapä- v.tr.-y ‘to imitate, to make fun of by imitating; to mock’

yachapänacU- v.i. ‘to say the same thing at the same time’

yachayllapa

1. *adv.m.* ‘slowly’

2. *adv.m.* ‘carefully’

¹³This category indicates that the verb may have a complement subordinated with *-na*.

¹⁴This category indicates that the verb may have a complement subordinated with *-y*.

musyay-yachay s.n. ‘wisdom, knowledge’

musyaj-yachaj s.a. ‘wise person, intellectual’

2. *v.tr.* ‘to like (some food)’

yachacä- v.i. ‘to acclimate, to adjust to a new place or circumstance, to get used to’

So I maintain that

1. derived forms have multiple senses so lexical databases should provide a way to accommodate these, as must the means for rendering these databases, and
2. lexical databases must provide a mechanism for indicating the relationship between a derived form and the sense of the base from which it is derived, and it should be possible to display a derived form as a subentry under the appropriate sense of its base. (Of course, whether and how to represent this relationship will depend on many factors: the tastes of users, the possibilities and economic factors of the delivery medium, and so forth.)

The Coward and Grimes [4] approach can accommodate either of these possibilities, but not both within a same dictionary. And Shoebox provides no support for subentries.

Bell and Bird [3, section 4.1] discuss the issue of “recursion,” of having an entry within an entry. They say,

“It seems intuitively clear that a complete model of dictionaries and lexicons should not need to include recursion of entries. That is, while sub-entries certainly occur, a survey of some 75 dictionaries and lexicons showed no evidence of sub-sub-entries. At worst, we have such entries as in the Quechua dictionary [Weber et al. [9] —DJW], where *manca-pantalun* is a sub-entry of *pantalun* with exactly the same format as its parent.”

But—as shown above—the Quechua dictionary surveyed does indeed have sub-sub-entries.¹⁵ In fact, it has sub-sub-sub-entries, but these are not readily apparent because (for this edition) we limited the display of embedding to two levels. For example, consider the chain of embedding *acä- > acaj > acajyä- > acajyaycachi-*. Although *acajyaycachi-* is a sub-sub-sub-entry, it is displayed at the second level on a par with *acajyä-*:

acä- v. ‘to be hot (ambient); to feel the sensation of having one’s hair pulled’

acaj s. ‘hot’

acajyä- v.i. ‘to swell and be hot (e.g., sprained ankle)’

acajyaycachi- v.tr. ‘to sprain’

So we must question the intuition expressed by Bell and Bird that “a complete model of dictionaries and lexicons should not need to include recursion of entries.”

To accommodate entries like *pantalun*, with its subentry *manca-pantalun*, the Document Type Definition given by Bell and Bird [3] does define Lexeme recursively. However it allows subentries only at the end of an entry.¹⁶ It does not permit a subentry under a particular sense, so does not accommodate many of the subentries in the Huallaga Quechua dictionary.

¹⁵Bell and Bird failed to see sub-sub-entries in Weber et al. [9] because they did not occur on the pages they chose to study (Steven Bird, personal communication).

¹⁶Their DTD for a lexical entry begins `<!ELEMENT Lexeme(Head, (Body|SimpleBody), Lexeme?) >`.

Bell and Bird [3, footnote 7] note that English *antidisestablishmentarianism* ‘opposition to the withdrawal of state support or recognition from an established church, esp. the Anglican Church in 19th-century England’ could be presented as nested subentries:

establish: to set up...
 establishment: something which has been set up...
 disestablishment: ...
 disestablishmentarian: ...
 disestablishmentarianism: ...
 antidisestablishmentarianism: ...

This displays the word’s successive derivation in terms of form, but is less than enlightening in terms of its meaning. But consider integrating this with multiple senses. *The Oxford American Dictionary of Current English* [2, 265] gives the following senses and sub-senses(!) for *establishment*:

- 1 the act or an instance of establishing.
 - 2 a business organization or public institution.
 - b a place of business.
 - c a residence.
 - 3 a the staff or equipment of an organization.
 - b a household.
 - 4 any organization permanently maintained.
 - 5 a church system organized by law.
- ...and so forth

It seems most appropriate to locate *disestablishment* under the fifth sense of *establishment*. Then, taking *dis-* in the sense of ‘removal of’ (as in *dismember*) the derived form naturally means ‘the removal of a church system established by law’, *disestablishmentarian*, ‘one who favors the removal of...’, and so forth.

This nesting of sense and subentries represents lexical information. One way to see this is to consider alternatives that give different—perhaps wrong—information. For example, locating *disestablishment* under another sense of *establishment* leads one to expect the derived form to mean something quite different. Or consider a different order of nesting. Suppose that *dis-* is first prefixed to *establish* to form *disestablish* and then *-ment* is suffixed to form (the noun) *disestablishment*:

establish: to set up...
 disestablish: to do the reverse of setting up, hence to dislocate...
 disestablishment: the act of reversing the setting up of
 and so forth

Is this sort of lexical information—a word’s successive derivation in terms of form—potentially significant (interesting, useful, ...) to users? I think so. For example, I think I can better wrap my mind around *antidisestablishmentarianism* for having understood it as a derivative of the fifth sense of *establishment*. And I think this sort of information is even more significant in languages in which derivation plays a greater role in structuring the lexicon than it does in English.

Therefore standards for lexical markup should facilitate the incorporation of this sort of information in the lexical database. One possibility is to allow a derived form to include a

reference to the sense of its base. Derived forms would then be entries in the lexical database rather subentries. And the form in which this information is displayed (if at all) would be left to the rendering process. To display them as nested under the senses would require that the rendering process recursively relocate and nest the entries of derived forms under the referenced senses.

For the Huallaga Quechua dictionary we took a different tack: each derived form was entered directly as a subentry under the appropriate sense of its base. This made for more complicated entries, but facilitated displaying derived forms as nested under senses, and was easier than incorporating the sense-derivative relationship with cross-references.¹⁷

That is, to support the nested rendering we incorporated the complexity in the database rather than in the rendering process. But—more significantly—by having the derived forms embedded under senses, they were localized to the entry of the base, making them more apparent and manageable.

Is the use of nested subentries the best way to present derived forms and the relationship to the sense from which they are derived? This will depend on many factors: the tastes of users, the possibilities afforded by the medium of delivery, and, of course, economic considerations.

Bottom line: lexical databases should accommodate (1) derived forms having multiple senses and (2) derived forms embedded under the senses of the bases from which they are derived.

Permit me a final reflection: Embedding derived forms under senses strikes me as so natural that I wonder why it has not become normal practice. The answer, I think, is that lexicography and dictionary making have adjusted their praxis to the limitations of small pieces of paper, to those precious 3×5 s that we file in shoeboxes. Until recently it was simply impractical to manage the complexity of embedding derived forms. We are now free from the tyranny of little papers, but on the verge of adopting standards that will perpetuate the limitations bequeathed us by those little papers!¹⁸

3 Some possibilities for future enrichment

Weber et al. [9] barely scratches the surface of the Quechua lexicon. Here are some ways its coverage and depth could be expanded (but for which there is currently no funding).

3.1 Audiovisual enhancement

We would like to enrich the database with sound and graphics, delivering the enriched dictionary on the web and CD-ROM (or whatever high-density portable device becomes available to potential users).

Sound would be used primarily to give voice to the Quechua expressions, so that a reader could click on a Quechua word or phrase and hear it. This could also be done for Spanish

¹⁷By the way, whereas it could be challenging to recursively relocate and nest the entries of derived forms under referenced senses, it is quite trivial to go the other direction, to convert embedded subentries into full entries for a “flat” display.

¹⁸For example, I believe it would not be possible to handle derived forms as subentries under senses within the standard being drafted for “Presentation/Representation of entries in dictionaries” ISO TC 37/SC 2 N258rev. of 7-MAR-2002.

and English.¹⁹

We would like to incorporate pictures of objects, particularly those hard to describe or identify (such as plant varieties). And we might include video clips of actions, cultural practices or events.

3.2 Automate the exploration of derived forms

Early in my fieldwork in Huanuco I had a surprising experience. To expand my knowledge of Quechua verbs I designed an 8.5×14 questionnaire. At the top was a blank in which to write a verb root. Down the left edge were several dozen combinations of suffixes that would form a word from a verb root, most combinations beginning with a derivational suffix. This was followed by a box in which a respondent could indicate if the word formed by the root and suffixes was acceptable. Following this was a line on which the meaning of acceptable words could be written.

I mimeographed a few hundred forms and wrote in roots. I then employed Teodoro Cayco V. to work at filling them out. However, after one week I abandoned the project. Why? Because it did not produce results? No! ...because it produced so much information that there was no way I could deal with it.

This approach should be attempted again. Now, however, rather than working on paper forms the process could be automated in such a way that the generated words would be displayed on a computer monitor and the human responses recorded directly in the lexical database.

3.3 Toward a Quechua ontology

My Quechua co-authors, Teodoro Cayco V. and Félix Cayco Z., once expressed curiosity about a thesaurus. (They had seen an English copy of *Roget's*.) I explained it to them briefly. They were already familiar with the concept of semantic domains as the result of some componential analysis we had done; see Weber [8]. When Felix, referring to the thesaurus, said, "Maybe we should make one of these," I suggested they begin to write down everything that exists, organizing it in a way that seemed fitting to them. I started them off by asking, "Well, what sorts of things are there?"

In less than two hours Teodoro and Felix made amazing progress toward a Quechua ontology. Here is a bit of what they wrote:

Cawajcuna 'what lives':

⋮

Animal 'animal':

Äbicuna 'birds'

Tacsha caj 'small': yuquish,...

Jatun caj 'large':

Aycha micoj 'carnivorous': anca,...

Mana aycha micoj 'not carnivorous': wachwa, gocha pätu,...

Chucaru 'wild':

¹⁹Some Quechua people would be keen to have the English. A rural school teacher once told me that he had Quechua-speaking students who wished to learn English, but they did not want to have to learn Spanish to learn English.

Jirca animal 'of the mountains': atoj, ...
 Munticho tiyaj caj 'of the rain forest': achu, gopi
 Chucaru uywa 'domesticated but by nature wild': wäca, cawallu, ...
 Manshu 'tame':
 Michina caj 'to be pastured': wäca, ...
 Wasi uywa 'raised inside': jaca, ...
 :
 Mana cawajcuna 'what is not alive'
 :
 Mana charina caj 'what can not be grasped':
 Pucutay 'clouds': chaqui pucutay, ...
 Wayra 'wind': shucucuy, alli wayra, ...
 Yacu 'precipitation': gasapa, runtu, ...
 Tamyä 'rain': alli tamyä, löcu tamyä
 Chirapa 'drizzle': gori chirapa, ...

This brief exercise turned up words and phrases not found in the dictionary and exposed facets of how the Quechua people categorize the world (witness the distinction between animals that graze and those that are raised inside).

Hopefully this enterprise can some day be pursued seriously, encompassing all the words currently in the dictionary, incorporating into the dictionary all the words and phrases it turns up, refining the categories, and so forth.

The natural habitat for the resulting ontology is the lexical database. However, I doubt the wisdom of trying to incorporate it within the current approach of a record-structured list of attribute-value pairs. Fortunately, more sophisticated approaches are becoming operational and I will eventually adopt one of these for the Huallaga Quechua dictionary.

4 A parting thought: the implications for endangered languages

To really document the lexicons of the world's endangered languages will require engaging people like Teodoro and Felix. A potential obstacle is the difficulty of the technology each must learn to contribute effectively. If the learning curve is too steep, if too much lexicography is required, if program interfaces are too complex, ... it will be difficult to engage people with little computational experience and formal education. So let's take care lest we create standards, computer programs, interfaces, ... that narrow the circle of potential contributors.

Given today's computation and interconnectedness we need to reconsider what dictionaries should now be and the impact that they might have. For example, it should be possible for a user to view the dictionary in different ways: different types of organization (alphabetical, ontological, with derived forms under the sense of their base or as separate entries), the ability to toggle on/off different types of information, and so forth.

And we should be alert to the potential impact of engaging a wide range of speakers in the creation, maintenance, expansion, ... of THEIR dictionary. In some situations this could contribute significantly to the stabilization of an endangered language. For none stand in greater awe of a language than those who have grappled with its lexical complexities.

References

- [1] (author not given). 2000. *The Linguist's Shoebox: Tutorial and User's Guide*. Waxhaw, North Carolina: SIL International.
- [2] Abate, Frank R. (editor in chief). 1999. *The Oxford American Dictionary of Current English*. New York: Oxford University Press.
- [3] Bell, John and Steven Bird. 2000. "A Preliminary Study of the Structure of Lexicon Entries" paper presented at the workshop on Web-Based Language Documentation and Description. 12-15 December 2000, Philadelphia, USA.
- [4] Coward, David F. and Charles E. Grimes. 1995. *Making dictionaries: A guide to lexicography and the Multi-Dictionary Formatter*. Waxhaw, North Carolina: Summer Institute of Linguistics.
- [5] England, Nora. 1995. "Doing Mayan linguistics in Guatemala". *Language* 68:29–35.
- [6] Herrero S.J., Joaquín and Federico Sánchez de Lozada. 1979, 1983. *Diccionario quechua: estructura semántica del quechua cochabambino contemporáneo*. Cochabamba: C.E.F.CO.
- [7] Poser, Bill. 2000. "A Sorting Tool and Issues in Sorting" presented to the Workshop on Web-Based Language Documentation and Description, Philadelphia, December 2000.
- [8] Weber, David J. 2000 "Lexicografía quechua: Desafíos y posibilidades" in Miranda, Luiz (editor) *Actas del I Congreso de Lenguas Indígenas de Sudamérica* 41–60. Lima: Universidad Ricardo Palma.
- [9] Weber, David John, Félix Cayco Zambrano, Teodoro Cayco Villar, and Marlene Ballena Dávila. 1998. *Rimaycuna: Quechua de Huánuco*. Serie Lingüística Peruana 48. Lima: Instituto Lingüístico de Verano.