Yalalag Zapotec Trilingual Dictionary
Zapotec-English-Spanish[1]

Heriberto Avelino

If a word in the dictionary were
misspeled, how would we know?
Steven Wright

0. Preliminaries. The *Yalalag Zapotec Trilingual Dictionary. Zapotec-English-Spanish.* started in Los Angeles, California, as a side project of my dissertation (2004) and out of the the collaboration with a key language consultant of and collaborator Jose Bollo, who is listed as co-author of the dictionary. This document presents the profile of the dictionary, and the reasons that guided the decisions in its design. Whether the dictionary can be taken to a greater scale depends on a number of variables out of the control of both authors; This document reports the progress reached so far.

So far, the database of the dictionary contains about 1000 entries. Close to half of these are nouns, 38 correspond to adjective-like categories, 47 registers of verbs, and 68 of other categories (conjunctions, negation, question words, numerals, among others). The database has been stored in the *Shoebox* program (SIL). José Bollo is responsible for the Zapotec part, while the glossing in English and Spanish has been almost done by me (native speaker of Spanish); it could be desirable to have the proof-reading of a native English speaker, since the dictionary is intended to be trilingual.

In making this project I admit overtly the influence of other 'senior' and professional dictionary makers. I have studied the Zapotec dictionary by Munro & Lopez (1999), which represents one the most thoughtful contemporary dictionaries on Zapotec languages. I also have been benefited from the manual of Bartholomew and Schoenhals (1983), a treatise which certainly have the advantage of synthesizing the experience of many linguists faced with the particular problems found in Mexican Indian languages. Other general studies on lexicography which have been important to this project, though its presence could not be detected at first sight, are Zgusta *et al* (1971) and Landau (1984). Very fruitful and direct lessons came from the discussion with the members of the 'dictionary making' class tought by Munro in 2002, to whom I thank here.

Notwithstanding the admited influence of the authors mentioned above, The *Yalalag Zapotec Trilingual Dictionary* project (henceforth quoted indistinctly also as the *Yalalag Project*) departs from them in some respects which are reflected in the overal design of the entries as decribed later. For example, any of the two most recents dictionaries on Zapotec languages Munro and Lopez (1999) and Butler (1999) include Zapotec etymologies.

---

[1] Yalalag Zapotec is an Otomanguean language spoken in Yalalag, municipality of Villa Alta, Oaxaca, by 3,500 in Mexico (2005) and 2,000 in Yalálag according to Ethnologue. Based on non-controlled observation of informal conversation among speakers, I was able to confirm that Yalalag is inteligible at some extent with the languages of other towns of the Sierra, particularly Yatzachi and Cajonos Zapotec.

Perhaps because there is not yet any complete list of Zapotec protoforms, o because there are various competing proposals (Fernández de Miranda 1995, Benton, Kaufman) it seems premature to include them as part of the article; however I believe that it is better to include the information available. In the present project I have included the protoforms as proposed by Fernández de Miranda (1995), the work that is accessible to me; however, I have considered later on including also the other reconstructions. In the same way that the English speaker is curious in reading the Proto-Indoeuropean roots of his contemporary language, Zapotec speakers are very intrigued in knowing that their language had a 'mother tongue'. Even though the reconstruction of Proto-Zapotec is incomplete, always it is possible that a larger reconstruction list comes out.

Other discrepancies will be evident by looking at the entire design of the *Yalalag Zapotec Trilingual Dictionary* project as discussed in the sections below. I will justify the decisions taken with respect to the style and content; first, I present the issues related to the orthography used in the dictionary such as the tension between the phonological analysis and the spelling system; the practical needs of the audience and the election of a set symbols appropriate. Then, I discuss the rationale of the design of the entries; the election of the citation form and the equivalents in the base languages; the justification for sample sentences and explanatory and technical sections in the record. Finally, I include the body of the dictionary compiled so far. The present stage of the project is ongoing, undoubtedly it should be modified and polished, the criticism to the present draft could be highly significant before of runing a test with the final users of the Yalalag dictionary.

1. Audience
The main intended audience of the Yalalag Zapotec Trilingual Dictionary (henceforth YZTD) is the native speakers of Yalalag Zapotec, particularly those living in USA; notwithstanding, it does not mean that other public is deliberately excluded; on the contrary, the present project has the goal of serving to other potential audiences. For instance, José Bollo, the main native speaker collaborator in this project, has pointed out that young people whose parents were Zapotec immigrants have vaying degrees of competence in the language of their parents. They are bilingual in English-Spanish and some of them also speak Zapotec. José Bollo thinks that for those children who technically are not native speakers of Zapotec a dictionary could be an incentive for learning the former language of their family. YZTD will be also of some service for linguists; this aim is reflected in several parts included in the design of the entries, especially grammatical information like part of speech, morphological analysis, etymology and verb classes among others.

In sum, the preliminary design of the entries as discussed here acknowledges explicitly the diversity of the audience and the priority of the native speakers settled in United States. At different degrees, YZTD will try to satisfy the needs of others interested in it, namely linguists, and more broadly, people interested in Zapotec languages.

1. Orthography and Phonological analysis
One of the most interesting issues in developing the Yalalag project was the preparation of the spelling system. Necessarily, the writing system entailed a preliminary phonological

analysis such that it represents (at least) the phonemic contrast in the language. In the initial stages, the task was not easy since I worked out both the phonological analysis and the orthography simultaneusly; thus, very often a modification on the view of the sound pattern was reflected in changes in the spelling system. Although there are still theoretical phonological problems, the core of the sound pattern has been understood, and the orthography is now systematic. In this section I would like to present some of the thoughts, problems and decisions arisen in the making of the Yalalag project.

In making a dictionary, the design of an orthography entails more than assigning coventional letters to a given language. Almost any manual on lexicography assumes the writing system of a language as a given fact, so that the task of developing an orthography is no very often contemplated as an assigment to the lexicographer. Neither Zgusta *et al* (1971) nor Al-Kasimi (1983) allude to the issue, and Landau mentions explicitly "...before a dictionary can be written for a language, the language must have developed more or less standard spellings" (1984:76). Since this could not be a topic for languages with a well-know history in writing, clearly this is no the case in the practice of making dictionaries for languages who lacks of a written tradition.

Indeed, choosing the spelling system is an unavoidable assignment for most linguists involved in these kind of projects, particularly true for those working on Amerindian languages. In general, it seems that a phonemic oriented (one phoneme-one symbol) orthography is a desirable feature in the dictionary. For example, Smalley (1964 *apud* Benton 1999) offers one of the earliest thoughts on this respect. He lists five principles for a succesful orthography: i. Maximum motivation, ii. Maximum representation of phonemic contrast, iii. Maximum ease of learning, iv. Maximum transfer into the national language, v. Maximum ease of reproduction. Of these, only point (ii) refers to a strict technical linguistic issue, namely the phonemic representation in the spelling system.

One of the 'two principles' of Munro (2000) claims explicitly "Each letter o letter combination in the system should represent the same sound each time is used, according to the pronunciation rules of the language". The position offered by Munro is interesting: the phrase 'pronunciation rules' allows us to have allophonic spelling and even multiple associations such that one symbol could have more than one pronunciation.

In the current project we have followed, as close as possible, the 'phonemic principle' as guideline in the orthography. In principle, each sound of the language is represented by the same symbol. Still and all, due to frequent allphonic and productive morphophonemic processes, we have faced problems identifying underlying and surface forms such that the spelling of several words were corrected when the rule was discovered. However, before continuing, I will introduce Yalalag Zapotec phonology and the issues of how to represent the sound pattern in a practical orthography.

1.1 Segmental Phonology. According to my analysis Yalalag Zapotec (henceforthYZ) has 22 consonants and twelve vowels (four modal vowels plus four checked and four rearticulated vowels). Furthermore YZ has contrastive tone, so far I have identified three tone patterns High, Low and Falling tones. The phonemic inventory of YZ is outlined below.

The consonants include the following phonemes:

voiceless obstruents	/ p, t, t͡ʃ, k, kʷ, ʁ, s, ʂ/
voiced obstruents	/ b, d, g, z, ẓ, ʒ /
approximants	/j, w/
nasals	/m, n, n*/[2]
laterals	/l, lˀ/
trill	/r/

And the vowels:

modal vowels	/a, e, i, o/
rearticulated	/a'a, e'e, i'i, o'o/
checked	/a', e', i', o'/

There are two previous orthographic systems for Yalalag Zapotec: López and Newberg (1990), and Newberg (2000, personal communication). The differences between those works and our proposal focus mainly on the symbols employed for:

a. the fortis-lenis contrast
b. the uvular
c. the labialized velars
d. tone specification

The table below summarizes the three versions discussed here (the shaded cells indicate that there is no orthographic equivalent). The column indicating phonemes in fact summarizes the phonemes claimed by the three proposals (which indeed keep some differences). For example, Newberg claimed that /gʷ/ and /ʁʷ/ are phonemes. As long as I do not have evidence supporting that claim, the alleged segments do not have a representation in our orthography. Conversely, I propose a phoneme /w/ which is absent in the other accounts, in consequence, the only column with a symbol for that sound is under Avelino & Bollo.

Phonemes and orthographic correspondences.

Yalalag Phonemes	Lopez & Newberg 1990	Newberg 2000 (p.c.)	Avelino & Bollo (in progress)
p	p	p	p
t	t	t	t
t͡ʃ	ch	ch	ch
k	k	k	k
kʷ	kü	kw (?)	kw

---

[2] Following the current convention used in Korean, the superscript asterisk marks fortis segments.

4

b	b	b	b
d	d	d	d
d͡ʒ	<u>ll</u>	ll	
g	g	g	g
gʷ		w	
ʁ	g̈	j	j
ʁʷ		jw (?)	
s	s	s	s
ṣ	x	x	x
z	z	z	z
z̦	<u>x</u>	xh	xh
			sh
ʒ	ll	ll	ll
l˙	l	l	l
l	<u>l</u>	r	r
j	y	y	y
w			w
m	m	m	m
n˙	n	nn	nn
n	<u>n</u>	n	n
r	r		rr
a, e, i, o	a, e, i, o	a, e, i, o	a, e, i, o
High, Low Falling, Mid tones			´ = high, ` = low ^ = falling

In fact the differences between Newberg's orthography and our proposal are well identified. Basically, the discrepancies are related to the phonemic analysis, i.e. the recognition of some units in one account but not in the other, or the election of different symbols for the same phoneme, as the ones already noticed. Let us consider the more salient divergences.

Newberg uses *ll* for both sounds d͡ʒ and ʒ. Recently we have found that there is no affricate in the speech of José Bollo but only /ʒ/. This approach is in consonance with the description by Newberg, who merges the two phonemes in one. Accordingly with his report, there is neutralization between those units "/j/ occurs only in the speech of some Yal[alag] Z[apotec] speakers. It has neutralized with /ž/ in the speech of others" (1983:5). Further research is necessary in order to determine the status of the alternation noted by Newberg. As for the present project, we will represent /ʒ/ with *ll*.

Another difference in the analysis is that we have registered [ʃ], represented by the digraph *sh,* an allophone of *ch* and *ll.* In fact, Newberg does not consider [ʃ] to be a phoneme.

Although it is true that there are no many lexical minimal pairs containing this sound, there are some occurrences as result of morphophonemic alternance. Whether or not [ʃ] is a phoneme (with crystal-clear minimal pairs) it has a function in a contrastive relation so that I have decided that the sound should be represented as a separate symbol in the spelling.
(1)

ʃ	shá'	'I am angry'	ʒ	llá'	'I will be angry'
	sháò'	'you are angry'		lláò'	'you will be angry'
	shé'	'he is angry'		llé'	'he will be angry'
ʃ	shuá'	'I am coughing'	tʃ	choá'	'I will cough'
	shúiò'	'you are coughing'		chúiò'	'you will cough'
	shué'	'he is coughing'		chué'	'he will cough'

The representation of the voiced sibilant /ʒ/ presents an interesting question. In principle the digraph *ll* could trigger a potential ambiguity in the sequence of *l-ll* i.e. [lʒ], as in the word *malllit* 'damn'. This is a latent problem since the sequence [ʒl] is quite possible if the initial consonant a verb stem is [l] and it is prefixed by the habitual marker *ll*. We do not have a solution based excusively on ortographic conventions. If such cases would appear, may be we could clarify the ambiguity by breaking up the morpheme boundaries.

Newberg does not list /w/ as part of the phonemic inventory of YZ. My analysis does include it as a phoneme. Apart from the phonetic description, one piece of the argument comes from syllable structure. So far, we do not have onsetless syllables in native YZ words. If /w/ were analyzed as a vowel, we could have highly suspicious open syllables and peculiar sequences of vowels. Consider the words below.
(2)

witj	'play'
wì	'orange'
waye	'firewood'

Another piece of evidence comes from morphology. Newberg reports an alternance between *u- ~ b-* 'Preterito' which I interpret as COMPLETIVE, *w- ~ b-*. Indeed, we have confirmed that for some verbs there is a free variation of both forms. Lastly, this alternance would suggest that the phoneme should be [- syllabic], in a standar phonological representation].
(3)

wllelle' ~ bllelle'	'he cried'
wlle' ~ blle'	'he pulled'

Moreover, the native intuition on syllabification favors the interpretation as a glide. José said that the words in (4) contain two and one syllables, respectively; this suggests that *w* does not count as syllabic.

(4)

wxwabe'	'he put'	COMPLETIVE
wnie'	'he '	COMPLETIVE

Besides of the analytical differences, both orthographies, Newberg's and our proposal are able to be reproduced in a conventional typewriter. Most scholars agree that a practical orthography should be easily able to type, for instance Bauernschmidt notes for American Indian languages in México "the symbols chosen should be available on the standard Spanish typewriter keyboard, printing presses or typesetter" (1977 *apud* Benton 1999:6); In the same sense, Munro points out as a principle that "an orthography should be typable on a standard typewriter or computer keyboard" (2000:1). The orthography used in the *Yalalag Project* tries to achieve this requierement, however, since we have decided to represent tones with the system of accents, that decision could cause some complications. Nevertheless, we think that tone is an important feature of the language, and that the audience intended in the project and linguists could be benefited by including tone in the entries.

1.2 Tone and stress. As indicated in the table above, we use the conventional system of accents over the vowels repeated here for convenience: ´ = high, ` = low, ^ = falling. Low is the 'unmarked' tone in the language. Very often in the dictionary there are words left without an expressed indication of tone; for those, I assume that they are specified as low tone. However, we write grave accent when there is an conspicuous minimal pair, or when the inflection has produced a derived low tone. Some of the pairs showing the evidence for contrastive tone are presented in (5) below.

(5)

zá	'fat'		zà	'bean'
lbá'	'throat'		lbà'	'gourd'
lé'é	'he, she' (resp.)		ré'è	'his name'
yé'èz	'pot'		yèz	'ear of corn'

I do not have a straightforward analysis for stress yet. Disyllabic words are the more common pattern in YZ, stress can be either on the ultimate or penultimate syllable. I believe that it is root-governed and can be predicted most of the time.

2. Article entry

Definitions for nouns and verbs will contain similar general parts. These are in essence information about the usage of the word, sample sentences explaining the use of a word in actual speech and diverse linguistic annotations, such as morphological analysis, clarification of underlying stems and affixes, variants in pronunciation among others. In this section I will cover the shared components for all the entries in the dictionary with exception of a. headwords, 2. part of speech, and 3. subentries, which will be discussed under the particular class of word.

1. *Usage.* As in any other language, YZ has a wide range of speech registers, conservative, vulgar, taboo, terminology restricted to specific groups or activities, etc. I will include this

information to the extent that I am confident of the appropriate use of the word as explained by my co-author. This entails that at some point the intuition of my collaborator should be confirmed with other speakers. Let us see an easy example. Proper names in many languages have affective forms (close related to baby talk) technically labeled 'hypocoristics'. In Yalalag Zapotec the use of hypocoristics in proper names is quite extended. We have agreed that this information is valuable for both, the main intended audience of the project, and for linguists.[3] Let us present some examples. The word for the Spanish name 'José' in YZ has two variants, the regular one, *Xhébé* used as vocative, and *Kwsé,* the form which conveys affection or closeness with the addresee. The second example shows a kinship term, 'sister', that is used exclusively by women. Finally, the word *mal-llit* is considered a vulgar word for zapotec speakers. We include usage information with the corresponding abbreviations between parenthesis after the part of speech and before the English equivalent.

> **Kwsé** *n* (*hypc*) Joseph | *José.*
> **Kwsén le miu nákbe'** | Joseph is a good friend | *José es un buen amigo. Span* José.

> **bire'** *n* (*fem. speech*) sister (speaking among women)|*hermana (hablando entre mujeres).*
> **le'e zu ito bibire'** | she has another sister | *ella tiene otra hermana.*

> **malllit** *adj* (*vulg.*) damn | *maldito.*
> **bèkò' mallliten bdauba' chixh'** | that damn dog, he ate my food | *ese perro maldito se comió mi comida. Span maldito.*

Another reason for including this field is that Zapotec languages make use of a highly developed system of pronouns differentiating various degrees of respect, sacred entities, sex, age and animals, among others. Particularly in Yalalag Zapotec there are four different forms of third person *lé'é* 'he' (respectful), *lebé'* 'he' (familiar), *lebá'* 'he' (animals), *lén* 'it' (things). We think that a comprehensive definition should contain this distinction.

> **lé'é** *pron (resp.)* he, she | *él, ella.*
> **le'e ba llbixe' rao yay** |he is falling down from the tree | *él se está cayendo del árbol.*

> **lebé'** *pron (conf.)* he, she| *él, ella.*

---

[3] The analysis of hypocoristics has been of considerable importance for recent theories in phonology (Itô, Kitagawa and Mester 1996).

lébeˊ' ba llbixbe' rao yay |he is falling down from the tree | *él se está cayendo del árbol.*

lebá' *pron (animal)* it (animal) | *él, ella (animal).* **lebá' ba llbixbá' rao yay** |he is falling down from the tree | *él se está cayendo del árbol.*

lén *pron (inanimate).* **lén ba llbixen rao yay** |it is falling down from the tree | *eso se está cayendo del árbol.*

2. *Illustrative sentences.* It seems that successful dictionaries contain sentences explaining by example the use of words. Bartholomew and Schoenhals (1983) discuss profusely in a chapter dedicated to that matter the reasoning of illustrative sentences in a dictionary. Among others, they claim that these sentences should 1. clarify reader's comprehension of the lexical entry, 2. clarify potential ambiguities (from the gloss in the base language), 3. help to improve the gloss.

It is a goal of this project to include in the definition at least one sentence (with the respective equivalents to English and Spanish) illustrating each headword; for cases where there is more than one sense or even for subentries it could be highly desirable to have sample sentences clarifying the usage and particular meaning of the word. To the extent that it is possible, I will include sentences coming from actual texts; however, others will be obtained directly from elicitation.

3. *The issue of anisomorphism.* If the phrase *tradutore tradittore* is the currency for translators, bilingual lexicographers are the biggest traitors, since they are 'translating' not just a closed text, but the entire lexicon (and by extension, the grammar) of one language into another. Even in a monolingual dictionary the perfect synonym is an idealized goal; thus, in a bilingual dictionary equivalents should be regarded with reserve inasmuch as two distinct linguistic systems are compared. Zgusta *et al* (1971) have labeled with the name of *anisomorphism* the problem of matching words and their meanings across languages. As for the Yalalag Project we have faced the anisomorphism issue at various levels, the most noticeable problem is when there is an item in YZ which is non-existent in the culture of the base languages. For example, ***yèl llìn*** is a tree and its fruit, in Spanish it is known as *chicozapote*; however, there is no name in English since the tree is not native from North America. These kind of problems multiply exponentially in a trilingual dictionary. For instance, YZ has the word ***jture'*** translated into Spanish by my collaborator as 'huajinijuil'. The word is uncommon to standard Spanish, and we do not have any clue of the equivalent in English. One solution adopted in several dictionaries in defining of species is the use of scientific terminology, so for the *chicozapote* word, the scientific identification *Achras zapota* (!) could help to some extent; nevertheless, given the characteristics of our target public this could not be a felicitous decision. Moreover, I do not qualify for making a

9

technical professional identification of animals or plants. A decision is still to be made on this regard.

The cases presented above illustrate the anisomorphism effect triggered by the abscence of culture specific items between the languages described in the dictionary. However, the notion can be applied beyond this context; Zgusta *et al* claim on this respect "Even in those areas where the two cultures overlap and where the material extralingustic world is identical, the lexical units of the two languages are not different labels appended to identical notions" (:296). For instance, one could think that body parts terminology tends to match across languages. This is a false expectation; let us see what occurs in YZ when compared to English. In YZ the word for 'hand' is *ná*, however the lexeme in YZ also designates the English 'arm' and 'forearm'. On that account *ná* could not be defined fairly as 'hand' and 'mano' in the dictionary, in any case a translation such as 'upper extremity' seems closer to the Yalalag Zapotec notion than the simple 'hand'. Regardless, we have decided to list the three different words that cover the meaning of Zapotec *ná*. This solution seems more transparent for the users-speakers of English and Spanish.

> ná *n* hand, arm, forearm
> (upper extremity) │ *mano, brazo,*
> *antebrazo (extremidad superior).*

Several types of dissimilarities are also found very often in verbs. For example, in Spanish (and may be in English) the verb *caer* 'fall down' is the verb used for describing the event of moving from up to down because the effect of inner weight and gravity. If more ellaboration is required in Spanish with respect to the specific configuration or direction of the 'falling down', the strategy is to specify, for instance, the body part that falls first *caer de cabeza* 'fall down head-first', *caer de espaldas* 'fall down bacwards'. However, YZ has three different lexical verbs indicating the specific 'style' of falling. The stem *bix-* refers to the act of falling down head-first from an elevation; *zà-* is the root used for falling backwards; and *xhix-* is the verb for falling down from being stood up. Our solution for these cases is to ellaborate on the definition using a paraphrase of the action described by the verb in YZ.

llbixe' *vi* falls down │ *(se) cae.*
  llbix yieren Gabriel falls down (being stood
  up) │ *Gabriel (se) cae (estando parado).*

llaze' *vi* falls down (backwards) │*(se) cae (de*
  espaldas). **llaze'** │ he is falling down
  (backwards) │*él se está cayendo (de*
  espaldas).

llxhixe' *vi* falls down (being stood up) │ *(se)*
  cae (estando parado). **llxhix yieren** │
  Gabriel falls down (being stood up) │ *Gabriel*
  (se) cae (estando parado).

4. *Linguistic information.* Under this label I include morphological analysis broadly, and in general all linguistic information about the headword, such as the independent elements in compounds, irregular forms of stems and other variants to the form as in the headword. Let us illustrate the encoding of linguistic information in compounds. The word for 'shrimp' in YZ is *bél daòˊ* which is composed by the word for 'fish' *bél* and *daòˊ* 'small', thus, literally means 'small fish'. The parts of the compound will be clarified in the section between curly brackets. The second example is similar in structure. As well as in previous case the meaning of the lexeme could not be recovered from the meaning of its parts 'water' and 'ancestors' which are spelled out in the linguistic information section.

> **bél daò'** *n* shrimp │ *camarón.*
> **bèl da'òn le lsa'ke zakba'**
> shrimp is expensive│ el *camarón está caro*
> {**bél** 'fish', 'pez' - **daò'** 'small', 'pequeño'}.

> **nis taò'** *n* sea │ *mar.* **nis taò'na'**
> **yielbana' le siaure**│ life if tastier
> in the sea │*en el mar la vida es más sabroza*
> {**nis** 'water', 'agua' - **taò'** 'ancestors',
> 'antepasados'}.

There is some important variation in the pronunciation of some phonemes in YZ, for example the alternation [u] ~[o]. Although we do not know if these differences are associated with extralinguistic variables, Newberg (1983) mentions some possible connection with a conservative-innovative speech distinction. At this moment we have decided just to register the variants, and later on we could specify further information. Variant pronunciations will be placed between braces, and introduced by the symbol '~'. Thus, in the example we are saying that 'sea food flesh' is pronounced either as **xho'** or **xhu'**. The form selected as the headword is the one prefered by José Bollo.

> **xho'** *n* sea food flesh │ *carne de marsico.*
> **wxe' gawa xho' bel** │ tomorrow
> I will eat fish flesh │ *mañana comeré carne*
> *de pescado.*{~**xhu'**}.

> **legake'** *pron* (*resp*) they (respectfully) │
> *ellos (de respeto).* **legake' ba**
> **llbixgake' rao yay** │they are falling
> down from the tree │ *ellos se están*
> *cayendo del árbol.*{~**leake'**}

5. *Protoform.* Personally, one of the features that I enjoy in any dictionary is the reference to etymology. My collaborator showed great curiosity regarding this part, and he believes that including the remote origin of words would be appealing to the community. Etymology is also of interest for professional linguists. Following his request, and the potential utility for the specialized public we have decided to include the protoform (Fernández de Miranda 1995) as part of the entry. If the word is a borrowing from other language, it also

will be specified in this place (for example *Span.* camisa, *Engl.* free way, *Nah.* chocolatl). Etymologies will be at the end of the entry. Protoforms will indicated by an asterisk.

> **bèkò'** *n* dog │
> *perro.* **bèkò' mallliten**
> **bdauba' chixh'** │that damn dog, he
> ate my food │ *ese perro maldito se*
> *comió mi comida.**beʔkuʔ .

> **klor** *adj* color │ *color.* **da klor zan ure**
> **nak xhaon** your shirt has many
> colors│*tu camisa tiene muchis colores.*
> *Span.* color.

6. *Subentries.* Typically subentries specify an extensions or ellaboration of the core meaning of the headword. Some dictionaries include as subentries categories such as idioms, methaporical expressions or secondary uses and meanings. We will use subentries in this standard form.

Thus, the primary meaning of ***xhàn*** is the body part 'buttocks', this (assumed) basic meaning is profusely extended to other figures and objects. So for instance, the low part of a hill or a house is referred with the word ***xhàn***. There is a similar relation in

xhàn llil 'base of the grill' referring to the part which makes contact directly with the fire.

So far, the meaning of ***xhàn*** can be figured out by the English and Spanish speakers; however, when the word also indicates the bottom of a page, the eye of a needle or the spine in a notebook, its clear that the meaning departs considerably between the languages of interest. We should indicate the different meanings in the base languages in the form of subentries. They will be in a separate paragraph below the definition of the main entry. Furthermore, in order to facilitate the identification, subentries will have an extra character indentation.

xhàn *n* buttocks; base; bottom │*nalgas;*
  base; punta; lomo de cuaderno. **xhàna'**│
  my buttocks ; *mis nalgas.*
   xhàn yia', xhàn yo', **xhàn nena'**│base
      of the hill, base of the house, first floor in
      a building │ *base del cerro, base de la*
      casa, primer piso de un edificio.
   xhàn cha', **xhàn llil** │ base of the pan,
      base of the grill │ *base de la cazuela,*
      base del comal.
   xhàn yich │ bottom of the page │ *pie de*
      página.
   xhàn laps, xhàn yiech │ the sharp end
      of a pencil, head, eye of a needle │ *punta*
      de lápiz, cabeza de la aguja.
   xhàn cuadern │spine of a notebook │
      lomo del cuaderno.

Headwords
Nouns
1. *Headword.* The headword uses as far as possible the morphologically simplest, unmarked form. For common nouns it is possible to use a similar form to the stem since there is no inflection for number or gender in Yalalag Zapotec; thus, headwords correspond to the singular, unpossesed form. There are a number of considerations for choosing this form, among the most important:

    i.   it is possible to pronounce the word in isolation
    ii. this is the first form obtained in elicitation
    iii. this is the form occurring with numerals

The word for 'snake' **bèlè** illustrates this use in the dictionary. Notice that the word in the sample sentence in (6) is **bèlèn**. In my analysis the ending **-n** functions like a determiner,[4] in complementary distribution with elements such as quantifiers **Xuán bétbe to bèlè / *bèlèn** 'John killed one snake'.
(6)
> **bèlè** *n*ˌsnake | *víbora.* **braù' bèlèn Xuan** | the snake bit John | *la víbora picó a Juan.* *beʔeLe.

For inalienable nouns (body parts and kinship terms, among others) I have chosen the form inflected for third person singular. The reason motivating this selection is that this form is the less complex morphologically, especially third person does not induce tone sandhi. There are three different suffixes marking possessor, as seen in the following paradigm *yichj -a'* 'my head', *yichj -o'* 'your head', *yichj -e'* 'his head'. Let us see a detailed description of this type of entries with the word equivalent to 'mother' *xne'.* First of all, it is no possible to have the root *na* in isolation, so that the word is obligatorily possessed. Accordingly, the word is a morphological complex form where *x-* is a mark of possession, *na-* is the root for 'mother', and lastly *-e'* expresses third person, singular; {x-na-e'} gives a full identification of morphological components. In general other further linguistic information is given between curly brackets after the sample sentence. Based on this illustration we can notice that the sequence *na-e'* produces the surface form *ne'* showing vowel deletion. I do not know how productive the rule is, so it seems reasonable to include the respective underlying morphemes. Secondly, since affixes are also entries in the dictionary —and it is planned to have a section explaining noun morphology— the reader can refer both to the definition of affixes and to the grammar section in order to obtain further information about the pieces mentioned in the morpheme by morpheme gloss of the headword.

> **xne'** *n* motherˌ | *madre.*
> **xne' llunè' yel'wao** | his mother is preparing the food |

---

[4] A sentence like the one in the entry example *brau' bélén Xuán* 'the snake bit John' rules out the possibility of *-n* as a case marker since the suffix is present in both nouns, *bélé* 'snake' and *Xuá* 'John'.

*su mamá está preparando la
comida.*
{**x-na-e'**} *na?a.

It should be mentioned that the form that my consultant gives first in elicitation is the correspondent to first person singular. Thus, our decision contravenes one of the advices of Bartholomew and Schoenhals regarding to the election of the best citation form "the form which speakers of the language regard as the natural citation form of that lexical item" (1983:31). Notwithstanding, the decision to include first person inflection in the headword could have further complications. In most zapotecan languages, as in YZ, there is a complicated system of 'tone perturbations' triggered by the first person singular. The consequences of preserving first person inflection would be that the headwords listed in the dictionary may be far from the unmarked form. In contrast, the third person singular possesor, in general, does not trigger tone changes.

Frequently one word in YZ has more than one meaning. For instance, in Mesoamerican languages the word for 'moon' and 'month' is often the same, as is the case for Yalalag Zapotec. The convention followed here is listing the multiple meanings in the respective base languages; for example, both equivalents 'moon', 'month', and 'luna' 'mes' will be entered respectively, in the English and Spanish sections, separated by a semicolon. Similarly, the word **yay** in YZ, as in several languages of Mesoamerica, refers to both, trees and firewood as illustrated in the second entry sample below.

> **bèo'** *n* moon; month | *luna; mes*
>   **na chixhjga' rue' beo'n za ito.**
> | I will pay you next month | *te
>   pago el mes que viene.*
>   **na yo' beo'xhen** | Today there is
>   full moon | *Hoy hay luna llena.*
>   *be?wu?.*

> **yay** *n* tree; firewood | *árbol; palo.* **le
>   bllia yay mankwon** How much the
>   mango tree produced now | *Como dio
>   el árbol de mango ahora!;* **na' ba tup
>   yay ka'** | Pick up these firewoods! |
>   *Junta esos palos!*

An important number of words in YZ consist of combinations of one or more words. However, the meaning of these words is unpredictable from the separate meaning of their parts. Compounds in the YZTD are translated, as far as possible, in parallel to the base languages. For instance, the word for 'hummingbird' in YZ is contains the root for 'suck' **go'x** COMPLETIVE and 'flower' **yaje**; in Spanish there is a transparent parallel compound 'chupaflor'; however, since the use of this word is not generalized to all Spanish dialects, I included also the more common word 'colibrí'. The following example, **nádon** 'lobster' illustrates a compound made of a noun and an adjective. Finally, 'ear-wax' is an example of a classical noun-noun compund in languages of the area. Clearly, in all this cases the lexical meaning cannot be derived from independent meaning of the parts, although it seems that

they are transparent for native speakers. In any case, for the purposes of the dictionary, the words involved in Zapotec compounds are always clarified in braces. Each word in isolation is given in the form as found in the entry, and then the gloss between single quotes in English and Spanish is reproduced.

> **go'xyajé** *n* hummingbird | *chupaflor, colibrí*. **go'xyajen ba' dau ris rakbá** the hummingbird is small | *el chupaflor está chiquito*. {**go'x** 'suck', 'chupar'- **yaje** 'flower', 'flor'}.

> **na'don** *n* lobster | *langosta*. **mèlen de to na'don ke be'** | Ismal has a lobster | *Ismael tiene una langosta*. {na' 'hand', 'mano' - don 'long', 'largo'}.

> **ye'nay** *n* ear-wax | *cerumen*. **bikuejo' ye'nayo'** Don't take out the ear-wax | *No te saques el cerumen*. {**ye'** 'excremento' 'excremento' - **nay** 'ear', 'oreja'}.

2. *Part of speech*. There is only one basic category for nouns in the YZTD (encoded in the entry as *n*). Subtle classification of this category is always possible (animacy, abstractness, individuation, comun, proper, etc.); however, incorporating this specific information in the dictionary in the part of the speech is out of the goals of this project. As for verbs, I have included just the basic valence relations transitive and intransitive. More on that subject will be discussed under the section dedicated to verbs. Other categories identified as part of speech are adjectives (*adj*), adverbs (*adv*), (*wh*) question particles, (*neg*) negative.

Verbs
1. *Headword. The issue on citation form and equivalents.* Different traditions on lexicography make different decisions about what is the citation form and the equivalents of verbs between target and base languages. The easiest case is when both languages match in inflection with the forms used as the headword and the correspondent equivalents. This is clearly exemplified in several Indo-European languages which use the infinitive form of verbs. Nevertheless, not every language has infinitive forms such that there is a suitable correspondence with each other.

This fact brings up the issue of what is the the optimal citation form in the target language, and what is the best equivalent in the base language(s). For instance, colonial dictionaries of languages of the Americas frequently used an inflected form as a lemma and translated into Spanish with the infinitive form. This was the model adopted by Molina (1571) for Nahuatl, and followed by others as Urbano (1605) for Otomi. Both authors take the first person singular, present in the target language, although the glosses in Spanish are always infinitive forms. Let me introduce one example below.
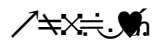
Tlatoa. ni. hablar algo, o chirriar, gorgear   'speak or , the birds sing'
     o cantar las aues. Pret. onitlato.

Hablar. nitlahtoua, nite, notza. Tana,yã.   'speak'
    ti,yã. tanatzôte. ti tzôte. Pl. hœ.

However, other traditions follow a complete parallel between the headword and the equivalent in the target language. A representative case is Modern Greek, a language lacking of infinitive form. The strategy in the *Greek-English Lexicon* presents the headword as the form inflected by first person singular in present, as well as the correspondent form in English. I reproduce a part of the definition for the verb 'to speak' ['omilo].

            ╱⊧⋇≒🍎 speak, converse, talk to...

Liddel and Scott (1961 [1843]:1222)

Several languages of the Americas do not have infinitives forms nor the verbal stem can constitute an independent word. On this respect, some of the dictionaries favor the parallelism between the main entry and the inflected form in the base language, instead of the infinitive form correspondent in English or Spanish, for example. This is the procedure followed in the collection of dictionaries by SIL, Munro and Lopez 1999, and others. Although it is possible to design a dictionary where the main entries are roots or stems, there is a tendency in avoiding this practice. For example, Hill describes the criteria for his Hopi dictionary "Entries that contain full information on inflection, definition and examples ("main entries") are made the simplest and most common forms words" (1998:xvii). Bartholomew and Schoenhals (1983) claim that the citation form should be accesible to native speakers, in their terms the citation form should be 'psycholgically natural' meaning that the headword is a word that the native speaker can pronounce in isolation, and it is the case tha bare stems or roots are often opaque for native speakers. Other dictionaries show alternative strategies, for example Munro and Lopez list verbs including the information of aspect-tense, but without the proper inflection for subject, such that the headword formally "cannot be used as full sentences" (1999:11); although the form in the headword lacks of subject, Munro and Lopez translated the form into English and Spanish using the fully conjugated forms correspondent to third person singular. Strictly speaking, this strategy fails in achieving the parallelism between the headword and the equivalents in English and Spanish whose third person forms are the subject of verbs.

As other Zapotec languages, Yalalag Zapotec does not have an equivalent to the infinitive form, and it is not possible to refer to the stem as an independent word.[5] Furthermore, inflected verbs have obligatory aspectual prefixes, this property makes difficult to sort the verb by the initial character of the stem.[6] Thus, we have decided that the form of citing

---

[5] There is a potential infinitive form in YZ. Verbs prefixed by *wa-* are translated by my collaborator as equivalents to Spanish infinitive; this is the form appearing in control verbs of the type 'I want to V'. I have decided investigate further this form before of accepting it as the YZ infinitive.

[6] In non-prefixal languages matching the order by the stem initial character is indeed possible. On this respect, the comment by Callahan is illustrative "I have been fortunate in my own research in that the Miwok

verbs is the habitual form, marked by the prefix *ll-* before the stem. My collaborator gives first and third person forms as the default cases. Accordingly to the former decision followed for nouns and considering the conditioned changes produced by the first person singular, I will use the third persons singular as the lemma in verb headwords.

> **llute'** *vt* kills | *mata.* **llute' ba'** |
>   he is killing an animal |*él está*
> *matando un animal.*

Although this procedure achieves a real word as the main entry, it has the detriment that most verbs are listed under a few letters in the dictionary. So, the potential relation between verbs and nouns (or any other category) is lost by the 'one pass eye' identification, which is indeed posible in languages where there no prefixes at all. Thus in order to get the stem one should ignore the prefix.

One this respect, it could be interesting to recall Molina's strategy. He succesfully accomplished the requirement of having an actual word as the headword, and furthermore he succeded in sorting the dictionary by initial stem in a language where prefix are obligatory in inflected words. However, such an achievement is paid in the morphological decomposition of the word, as long as the prefixes follow the stem in the headword (separated by a comma). Someone could find this mechanism disturbing for native speakers, however I could mention that I know native speakers of contemporary Nahuatl who can use Molina's dictionary efficiently after beig introduced to the basic structure of headwords. Perhaps in a near future we would explore Molina's system in our project.

Frequently, verbs have closely related words both semantically and by form. They are listed under the same entry if both semantics and shape match enough that they can be conceived as extensions of the core meaning of the basic form. For instance ***wa-e'*** means 'he carries'. This is the more general word in the sense that it describes a wide range of distinct events (where the prototypical action is carrying something on the back). A similar word ***wal-e'*** is translated as 'hold' or 'carry in the arms'. A relationship seems plausible, but it is not clear what is the morphological composition of the second one. Nevertheless, it is considered a specialization of the main entry, and in consequence listed as a subentry. Following the convention for subentries in nouns, we will take advantage of different degrees of indentation. Thus ***wale'*** is indented one character with respect to the body of the definition of the main entry.

> **llwé'è** *vt* carry (something on the back)
> | *cargar (algo en la*
> *espalda).* **llwé'è to da'**. He is
> carring a mat | *él está cargando un*
> *petate.*
>   **llwale'** *vt* hold (in the arm) |
>   *cargar (en el brazo).* **llwale'**
>   **toay**. he holds a towel on his arm

---

and Costanoan languages lack prefixes" (2000:333)

| *él carga una toalla en su brazo.*

2. *Argument structure.* Verbs are the more complex class of words in YZ. They convey information about valence and argument structure, as well as semantic-syntactic beahvior. In investigating verb entries for the present project I used as a guide the book *Verb classes and alternations in English,* by Beth Levin (1993).

Clearly it is not possible, even undesirable, to transpose the structure of one language to another blindly, hence that my intention in look at the verb classes in English was just with the aim of having a point of comparison in the behavior of YZ verbs.

Ideally, a dictionary should encode the idiosyncratic information of every lexical item, but also the information that enables a native speaker to use and predict the behavior of words; specifically with respect to verbs, I am thinking about the syntactic knowledge related to transitivity relations, particularities concerning to diatesis, and the selection of particular arguments which are encoded in the meaning of verbal lexical items.

Even though such a research program is quite attractive, it entails a refined knowledge of the language. In practical terms, for the present project the goal of representing as much as possible this subtle information has also represented frustration because of the slow pass of discovering the grammar behind each word. On the other hand, I have faced the dilemma of minimizing technical information in benefit of our audience and at the same time, expressing this important information; let me state straight this concern: even if for linguists this information is highly valuable, how useful is for a lay audience to know whether a verb is unacusative, or presents restrictions in middle voice, for example? We have decided that we will include in the part of speech only the basic distinction of transitivity, i.e. transitive (*vt*) - intransitive (*vi*) verbs. Thus, the essential argument structure would be minimally represented. Regardless, we have to figure out some way of encoding the other relevant information.

Verbs often encode semantic specifications beyond transitivity related to the specific nature of the arguments involved in the event described by the verb. Thus for example we have looked for verbs which select for specific subjects. We had some success in contrasting activities between humans and animals which were encoded as separated lexical items ('eat', 'walk', 'gave birth' verb types, among others), because YZ morphology has special morphemes specifying when the subject is an animal *-ba'*.

1 a. llaubá' yet	'an animal is eating tortilla'
b. llaubé' yet	'an man is eating tortilla'
2 a. bèkò'n za'bá'	'the dog is walking'
b.béne bièo zá'	'the man is walking'
3 a. llxhanbá' to mè' dào'	'an animal is giving birth to a calf'

b. llxhané' to bi dào'   'a woman is giving birth to a baby'

We can mention that we had relatively more success looking for verbs that differentiates the nature of their objects. For instance, the equivalent to 'take a shawer' in Yalalag Zapotec has a distinction split by the sex of the object. So, **_gazj ~ azj_** is the generic stem, the one used when the semantic object is male. Notice, however, that, as José Bollo has pointed out, this verb also could be used referring to a woman, as the examples illustrate below.

(7)

a. Yieri ll**azj**be'   'Gabriel is taking a shawer'

b. llw**azj** Pinna' bidao'   'Josephine bathing the boy'

c. Pinna' ll**azj**be'   'Josephine is taking a shawer'

However, there is another stem **_wch_** used exclusively for female objects. The ungrammaticality of (8b) accounts for the strong selection of this verb.

(8)

a. Pinna' ll**wch**be'   'Josephine is taking a shawer'

b. *Yieri ll**wch**be'   'Gabriel is taking a shawer'

c. Pinna' ll**wch**be' bidao'   'Josephine is bathing the boy'

Now let us see how can we encode the pertinent information clarifying the restriction on the object in the dictionary entry. It seems that the best solution is to spell out in the equivalents to the base languages the information concerning the more restricted verb, e.g. the one which is used exclusively for women. Since the restriction is about objects the indication will be in parenthesis after the verb in both English and Spanish (in the canonical position for objects in these languages).[7] In addition, the sentences in the entry could elucidate the meaning of the verb. The entries below illustrate the solution thought for this cases.

llazjbe' _vt_ takes a shawer│ _se baña_. **Yieri
  llazjbe'** │ Gabriel is taking a shawer │
  Gabriel se está bañando; **llwazj Pinna'
  bidao'** │ Josephine is bathing the boy │
  Josefina baña al niño.

llwchbe' _vt_ takes a shawer (a female) │ _se
  baña (una mujer)_. **Pinna' llwchbe'** │
  Josephine is taking a shawer _Josefina se está
  bañando._

3.*Idioms*. As compounds, idioms are idiosyncratic constructions which generate an unexpected meaning from the sum of their parts. For example in English 'to give a hand' is a phrase which really means 'to help'. We found that YZ has a parallel idiom, *we'na'be'* is

---

[7] The object of reflexives does not entails any problem for this definition.

literally 'he gives hand to' although it means 'he helps (to some one)'. For this cases we have listed the idiom as a subentry under both main entries, 'give' and 'hand'.

llwe' *vt* gives | *da.*
  llwe' yeo'n mell lachen Santiago is giving money to Lazaro *Santiago le está dando dinero a Lázaro.*
   llwena'be' *vt* gives a hand, helps | *da una mano, ayuda.* {**we'** 'give', *'da'* - **na'** 'hand' *'mano'*}.

na' *n* hand | *mano.*
  llwena'be' *vt* gives a hand, helps | *da una mano, ayuda.* {**we'** 'give', *'da'* - **na'** 'hand' *'mano'*}.

4. *Irregular verbs and paradigms.* Otomaguean languages are characterized for a large list of irregular verbs. Yalalag Zapotec is not an exception. In López and Newberg (1990) 33 out of 44 verb paradigms are labeled as irregular conjugations. At the present stage of the project we do not have a clear picture of the uniformity across paradigms, such that we cannot 'accuse' with certainty a verb of being irregular. Nevertheless, following the strategy of some other dictionaries we have have planned to indicate the conjugation of a given verb by indicating a model verb which shows a paradigm fully conjugated in an appendix. For example, based on López and Newberg (1990) we know that 'bite' **lli'ne'** is an irregular verb; comparing its inflection with that of 'swallow' **llebe'** we can stablish a parallelism, such that we know that it is also irregular, or better, that they are similarly irregular. The entry of 'swallow' in the dictionary could indicate the model of conjugation after sample sentences between parenthesis.

'bite' López and Newberg (1990)			'swallow'		
Future	chi'na'	1 Sg.	IRREALIS	cheba'	1 Sg.
	chi'no'	2 Sg.		chebo'	2 Sg.
	chi'ne'	3 Sg (fam)		chebe'	3 Sg.
Present	llyi'nllo	1 Pl.	HABITUAL	llebllo'	1 Pl.
Past	udi'nllo	1Pl.	COMPLETIVE	bdebllo'	1 Pl.

chebe' *n* swallows | *traga.* **chebe' yet** | he is swallowing tortilla | *el está tragando tortilla.* (*Parad.* chine').

Final note. The Yalalag Project, as any other dictionary, is in progress. I have discussed the preliminary stage in making our dictionary. We —the authors— are aware of some of the deficiences and gaps, but may be we do have not noticed several major problems. We are evaluating some of the decisions taken so far, specially we would like to reconsider the issue on the citation form and the equivalents in verbs; likewise, we are still thinking in how to represent fairly the syntactic and semantic content of verbs, which we believe should be

in the definition. I have no discussed in deep editorial aspects of the project, they should be revised at some point. I have relied on the Shoebox program for most of the technical aspects, however I should mention that the draft attached to this paper was modified substantially 'by hand'; the program saves time, but it does not produced the final product as designed and presented here.

I have been thinking in the question in the epigraph at the begining of the paper, I do not have an answer, but I wish that in our dictionaries we had less misspellings than we are aware.

References

Al-Kasimi, Ali M. 1977. *Linguistics and bilingual dictionaries.* E.J. Brill. Leiden.

Bartholomew, Doris A. and Louise C. Schoenhals. 1983. *Bilingual dictionaries for indigenous languages.* Summer Institute of Linguistics.

Bauernschmidt, Amy. 1977. "The ideal orthography", in *Nova Lit, 5, 3.* Pp. 1-8.

Benton, Joseph P. 1999. *How the Summer Institute of Linguistics has developed orthographies for indigenous languages of Mexico.* www.sil.org/americas/mexico/ilv/L001i-SILOrth.pdf

Callahan, Catherine A. 2000. "Writing a User-friendly dictionary", in Frwaley, Hill adn Munro (eds.) ms.

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.), 2013. *Ethnologue: Languages of the World, Seventeenth edition.* Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.)

Fernández de Miranda, María Teresa. 1995. *El protozapoteco.* Edición a cargo de Michael J. Piper y Doris A. Bartholomew. Serie Estudios de Lingüística y Literatura XXVIII. El Colegio de México, Instituto de Antropología e Historia.

Hass, Mary R. 1978. "What belongs in a bilingual dictionary?" in *Language, culture and history. Essays by Mary R. Hass.* Selected and introduced by Anwar S. Dill. Standford University Press, Standford, California.

Hill, Kenneth C. (director), Emmory Sekaquaptewa, Mary E. Black and Ekkehart Malotki (editors). 1998. *Hopi Dictionary. Hopiikwa Lavaytutuvemi. A Hopi-English Dictionary of the Third Mesa Dialect.* U. of Arizona Press. Tucson.

Itô, Junko, Yoshihisa Kitagawa, and R. Armin Mester. 1996. 'Prosodic faithfulness and correspondence: evidence from Japanese argot', *Journal of East Asian Linguist.* 5.

Hale, Kenneth and Danilo Salamanca. "Theoretical and universal implications of certain verbal entries in dictionaries of the Misumalpan languages" ms.

Landau, Sydney I. 1984. *Dictionaries. The art and craft of Lexicography.* Cambridge University Press.

Lara Ramos, Luis Fernando (director). 1996. *Diccionario del español usual en México.* El Colegio de México, Centro de Estudios Lingüísticos y Literarios. México.

Levin, Beth. 1993. *English verb classes and alternations. A preliminary investigation.* The University of Chicago Press. Chicago and London.

Liddell, Henry George and Robert Scott (comp.).1940 [1843]. *A Greek-English Lexicon.* Oxford University Press.

Munro, Pam. 2000. *Lexicography.* Course notes.

Molina, Fray Alonso de. 1571 [1992].*Vocabulario en lengua Castellana y Mexicana y Mexicana y Castellana.* Edición facsimile. Biblioteca Porrúa, 44. Editorial Porrúa. México.

Newberg, Ronald. 1983. *Yalalag Zapotec Phonology.* Manuscript.

Smalley, William A. 1964. "How shall I write this language?" *Orthography studies.* edited by William A. Smalley. Amsterdam, North-Holland Publishing Company. Pp. 31-52.

Urbano, Alonso. 1605 []. 1990. *Arte beve de la lengua otomí y vocabulario trilingüe español-nahuatl-otomí.* Edición de René Acuña. Instituto de Investigaciones Filológicas. Filología, gramáticas y diccionarios, 6. Universidad Nacional Autónoma de México.

Zgusta, Ladislav in cooperation with *et al.* 1971. *Manual of Lexicography.* Janua Linguarum Studia Memoriae Nicolai Van Wijk Dedicata. Series Maior 39. Academia, Publishing House of the Czechoslovak Aacdemy of Sciences. Mouton The Hague, Paris.