Appendix A: The MIG Zoque lexicon

A MIG Zoque lexicon with around 4500 entries can be accessed on-line at http://www.albany.edu/anthro/maldp/. This database was produced mainly by me, but significant numbers of entries were made by Terrence Kaufman, and all entries have been reviewed by him. The primary consultant for the lexicon was Germán Sánchez Morales, although most of the entries were reviewed with other speakers.

Lexicons for this and other MesoAmerican languages were developed by the MesoAmerican Languages Documentation Project using Shoebox 2.0 databases. They are gradually being made available as on-line databases; currently, those of MIG Zoque and Oluta Mixe (Roberto Zavala Maldonado) are accessible at the MALDP web site, and the Santa María Chimalapa Zoque lexicon (Terrence Kaufman) will appear soon.

An entry in the database is composed of many labelled fields, each of which contains a line of textual information about the lexeme. Many of these fields can be used as search criteria, in accordance with the instructions on the web page. Unfortunately, the system requires some prior knowledge of how the entries are composed and what sorts of information can be found in a given field; providing that information is the object of this appendix.

Table A.1 gives the complete list of field names and with descriptions of their contents. Not all of these fields are used in every lexical entry, and some of

them are peculiar to my lexicon. Entries are keyed by the *lexeme*, which is entered in its underlying form.

Field	Description
lex	The lexeme. The key field of the database.
PSH	The phonological shape (surface form) of the lexeme.
UND	The underlying morpheme breakdown of a polymorphemic lexeme.
MB	Morpheme-by-morpheme glosses (in Spanish) of the UND field.
VAR	Phonological variants of the lexeme.
GRAM	The grammatical code of the lexeme.
USE	Note about the usage of the lexeme; only used for bound morphemes.
SPG	Spanish translation of the lexeme.
ENG	English translation of the lexeme.
SPL	A supplemental form, used to determine the grammatical class of a verb. This will be one of: passive, antipassive, causative, assumptive, or perseverative.
SLGR	The morpheme gloss code of the supplemental form.
SLGL	Glosses of the supplemental form: Spanish // English
EXU	An example, in underlying morpheme-breakdown form.
XPSH	The 'phonological shape' of the example - written as it sounds.
XSP	Spanish translation of the example.
XEN	English translation of the example.
SYN	Synonymous entries.

SEMF	The semantic code for this root.
ASP	The aspectual class of the root.
ARG	The role of an argument that is specifically entailed by this root.
NOM	Nominalized forms of this root.
SUB	Subordinate entries - lexemes with this verb as root.
ETYM	Word in the donor language - used for loan words.
ETL	Language from which the lexeme was borrowed.
ROOT	The root verb for this lexeme.
RTGR	The grammatical code of the root: (verbs) T, I, or P.
DSRC	Initials of the linguist who collected the data and the year it was collected.
CMTS	Comments - information that didn't fit anywhere else.
DATE	Date that the entry was last modified.

Table A.1: Lexical database fields

The orthography used in the lexicons is different from that employed in this grammar. We use an ASCII-compatible orthography to simplify data entry, and to provide the speaker communities with orthographies that can be used on typewriters as well. The alphabet is listed at the top of the lexicon's web page, but some additional notes on the characters used are useful. Characters that may be unfamiliar are shown in table A.2. @ represents the 'sixth vowel' in the lexical databases. This is a high mid vowel in the other MZ languages, but a mid mid vowel in MIG Zoque. So, although I use ə for this vowel in the grammar, it will appear as ü in the pretty-printed lexical entries retrieved by the search.

MALDP	IPA
@	ə
7	ʔ
nh	ŋ
x	š
j	h
ch	č
tz	c

Table A.2: MALDP orthographic conventions

Lexical entries are written in their underlying forms; that is, the morpheme breaks are indicated by the appropriate morpheme break symbol. For example, the entry for ʔaŋkimmobáʔ, 'leader', is 7anh=kim.'oy.pa+a7k, essentially the morpheme breakdown of the word. The character ' is used to indicate a glottal

stop that gets deleted or that induces gemination of the preceding consonant, as is the case with the initial glottal stop of the antipassive suffix 7oy.

Similarly, W is used to indicate a /w/ that induces gemination in the preceding syllable. Such a W is found in the entry for ʔaŋmayyə́ʔk, 'student': 7anh=may-W@+V7k (MOUTH.count.COM+REL). Another symbol intended to provide information about historical connections among these languages is H, which indicates a glottal fricative that does not appear on the surface in some languages. These H's never appear on the surface in MIG Zoque, which is why I don't use them in the morpheme breakdown lines in my examples. An entry with this symbol is 7anh=kuk.'aH (ʔaŋkukka-, 'to collect').

These orthographic conventions are used in all the fields that represent the underlying form of the entry: lex, PHO, UND, VAR, and EXU (an example of the entry's use).

The next field that might of interest as a search term is GRAM, used for a grammatical code describing the entry. Some of these codes are shown in table A.3. The codes are based on Spanish words, and may be combined into strings reflecting the components of the lexical entry. For example, the code *sms* indicates a noun (sustantivo) modifying another noun. This is not a constrained set of items, and thus would be difficult to exploit in a fine-tuned search, but one could use it to find all the transitive verbs, for example.

Code	Spanish	English
vt	verbo transitivo	transitive verb
vi	verbo intransitivo	intransitive verb
s	sustantivo	noun
av	adverbo	adverb
a	adjectivo	adjective
pron	pronombre	pronoun
indef	indefinitivo	indefinite

Table A.3: Grammatical codes

Below is the entry for the verb root təŋ-, 'to cut with a machete'. The lexeme comes first, written as it appear in the lex field of the database. Next is the grammatical code: vt for *verbo transitivo*. Then comes the Spanish translation, followed by the English translation. Next will come the example sentences, if there are any. If there is more than one example, they will be numbered. Each example gets four fields: the underlying form (EXU), the surface form (XPSH), the Spanish translation (XSP) and the English translation (XEN). Supplemental forms follow the examples. These only appear in the entries for verb roots, and are the derived forms that were used in classifying those roots: the antipassive, passive, assumptive, perseverative, and sometimes causative forms. (Only the antipassive and the assumptive turned out to produce useful classifications, of T3 and P verbs, respectively.) Next, the root verb of the lexeme

is given, indicated with the symbol % and followed by its classification code in parentheses. The underlying form of the lexeme comes last, written between //.

> tünh **vt** cortar, con una machete o fierro // to cut, with a machete or iron tool. Ex: 7üy+ tünh.ten-Wü. [7üy tünhtennü]. Lo tiene cortado. // He had it all cut. Sup: 1 tünh.7oy %tünh (T3). Underlying form: //tünh//.

The semantic field (SEMF) may also be of interest for searches. Entries for the names of plants and animals use this field to indicate the ethno-biological group to which the lexeme belongs, such as ʔokoš, 'shrub', for short leafy plants. These items can be found by entering EZ or EB in the semantic field box. I have also used the SEMF field to make a rough lexical categorization of verb roots. The categories are listed in table A.4. Some roots may belong to more than one category; for example, caŋ-, 'to hit with the fist', is categorized as both *hit* and *hand*. ʔan-, 'to warm by the fire', belongs to all of the categories *fire*, *temp*, *cook*, and *body*.

animal	condition	light	spread
aspect	cook	move	temp
bite	corn	nature	touch
body	cover	perception	trade
break	emit	pick	water
build	fall	pick_up	
carry	farm	plant	
catch	fire	quantity	
change	hand	scrape	
clean	hit	sew	
cognitive	hunt	size	
color	keep	sound	

Table A.4: Semantic fields for verb roots