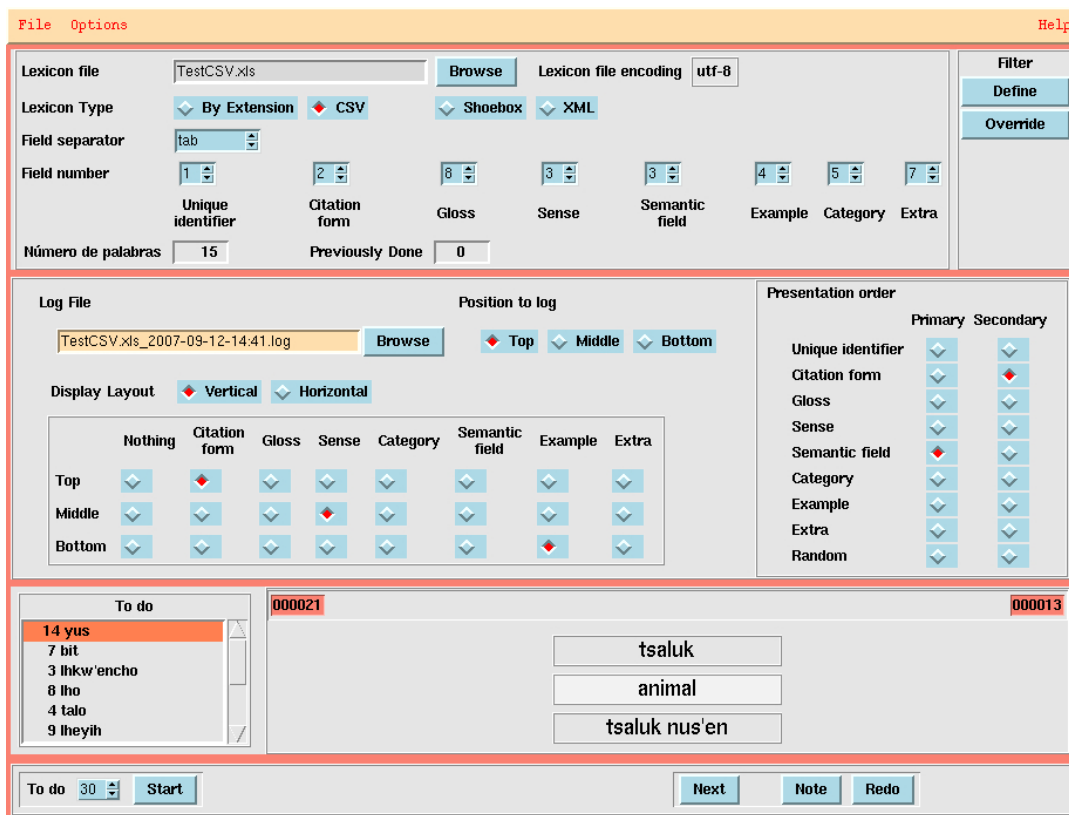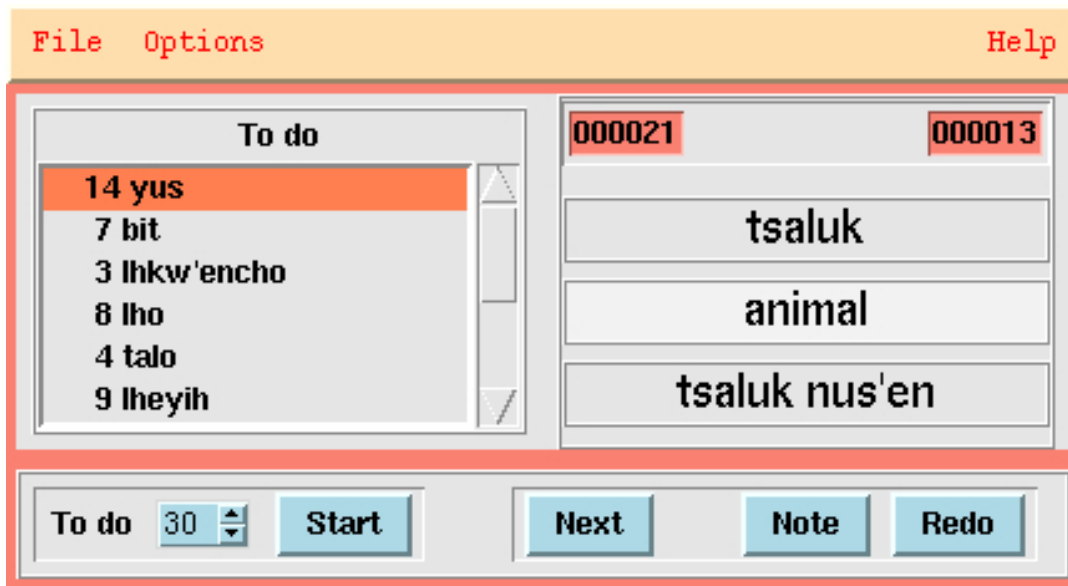# Prompter and Segmenter

Prompter and Segmenter are a pair of programs which together address the problem of recording large numbers of words and phrases to accompany a lexicon. Prompter is used to elicit the desired items, resulting in a single audio recording containing all items as well as a log file containing a time stamp for each item. Segmenter is then used to isolate each item in the recording and write it out in a separate, informatively named, file. Both programs are written in Tcl/Tk, a language available at no cost for all major platforms, including GNU/Linux, Mac OS X, and Microsoft Windows.

Prompter reads the items to be elicited from a lexicon, which may be in CSV, Shoebox, or XML format and applies user-specified filters to restrict attention to the desired subset of lexicon entries. It then presents one item at a time, holding the prompt until the user presses the "Next" button. From one to three components may be displayed for each item, e.g. the headword, an example sentence, and the lexical category of the word. The choice of fields to display and which field is associated with which position in the display is freely configurable. The order in which items are presented is determined by a sort on the user's choice of two keys, e.g. semantic field and citation form.

As seen above, the controls for the principal options are displayed by default when the

program starts up. Once the lexicon file has been loaded and the parameter settings are satisfactory, the controls may be hidden if so desired, as shown below.

File	Options		Help

To do

000021		000013

14 yus

tsaluk

7 bit

3 lhkw'encho

animal

8 lho

4 talo

tsaluk nus'en

9 lheyih

To do | 30 | Start | Next | Note | Redo

A variety of other options may be set from the Options menu. By default a lexicon is assumed to be in Unicode, but numerous other encodings are recognized.

ascii	big5	cp437	cp737	cp775	cp850	cp852
cp855	cp857	cp860	cp861	cp862	cp863	cp864
cp865	cp866	cp869	cp874	cp932	cp936	cp949
cp950	cp1250	cp1251	cp1252	cp1253	cp1254	cp1255
cp1256	cp1257	cp1258	dingbats	ebcdic	euc-cn	euc-jp
euc-kr	gb1988	gb2312	gb2312-raw	gb12345	iso2022	iso2022-jp
iso2022-kr	iso8859-1	iso8859-2	iso8859-3	iso8859-4	iso8859-5	iso8859-6
iso8859-7	iso8859-8	iso8859-9	iso8859-10	iso8859-13	iso8859-14	iso8859-15
iso8859-16	jis0201	jis0208	jis0212	koi8-r	koi8-u	ksc5601
macCentEuro	macCroatian	macCyrillic	macDingbats	macGreek	macIceland	macJapan
macRoman	macRomania	macThai	macTurkish	macUkraine	pascii	shiftjis
symbol	tis-620	ucs-2be	unicode	utf-8		

The interface can be made available in any language for which a message catalog is provided.

English/Ingles
Español de México/Spanish (Mexican)
Français/French
हिन्दी/Hindi
Português do Brasil/Portuguese (Brazilian)
Русский/Russian
中文/Chinese (Mainland))

Information about the timing of the recorded items is provided in two ways. First, a log entry is created for each item containing a timestamp with respect to the onset of the recording. Second, at the start of each batch a tone is emitted that will be picked up by the recorder.

Segmenter is a specialized audio editor designed especially for the efficient fragmentation of long recordings into small pieces. It provides multiple views of the audio waveform at different resolutions, the ability to move the window onto the audio waveform in a number of ways, automated setting of cut points at zero-crossings, and automatic generation of output file names or use of names read from a list. Segmenter identifies the synchronization tone emitted by Prompter and can read Prompter log files so as to obtain a list of the timestamps of the beginnings of each item.

Two of the mechanisms for positioning the window provided by Segmenter are of particular importance. First, the position of the left window edge can be set to a beginning of item timestamp. Second, the window may be positioned with respect to transitions between regions of sound and silence. (The buttons controlling such positioning are those in the bottom row, labelled "silence" in Chinese.) The regions selected may be written out to files directly or recorded in a list to be written out later, after further examination.

Segmenter is a derivative of SndBite, for which additional information, including an illustrated reference manual, is available at: http://billposer.org/Software/sndbite.html.