Appendix: Some useful regular expressions for ELAN users

1. Symbols

symbol	place	meaning
\b	at the beginning and/or the end of a string	word boundary
\w+	at the end of a string	variable end of word
.	anywhere	any letter
.*	between spaces	any string of letters between spaces/ any word
.*\	between spaces	any string of words
(x|y)	anywhere	either x or y
[^x]	place at the beginning	not x
(....)\1	anywhere	words with four reduplicated letters
?	after a letter	preceding letter is optional
(xyz)?	anywhere	the string xyz is optional

Table 1: symbols

2. Searching for particular complex word forms

symbols	hits	examples
sa	all words containing the string *sa*	*sa, vasaku, sahata, tisa*
\bsa	all words starting with *sa*	*sa, sahata, sana,* NOT *vasaku, tisa*
\bsa\b	all words *sa*	*sa*
\bsa..\b	all words consisting of *sa* and two letters that follow *sa*	*saka, saku, sana,*
\bsa\w+	all words beginning with *sa,* but not *sa* by itself	*sahata, sana*
\b.*ana\b	all words ending in *ana*	*sinana, tamuana, sana, bana, maana*
\b[^(bana|maana)].*ana\b	all words ending in *ana*, but not *bana* or *maana*	*sinana, tamuana, sana*
(....)\1	all words with four reduplicated letters	*pakupaku, vapakupaku, mahumahun, vamahumahun*
\b(....)\1	all words beginning with four reduplicated letters	*pakupaku* NOT: *vapakupaku*
\b(....)\1ana\b	all words beginning with four reduplicated letters and ending in *ana*	*vasuvasuana, hunuhunuana*
\bva(....)\1	all words with the prefix *va-* and four reduplicated letters	*vapakupaku, vagunagunaha*
\bvahaa?\b	all tokens of *vahaa* and *vaha*	*vahaa* and *vaha*

Table 2: Combinations of symbols on word level

3. Searching for particular sequences of words

	symbols	hits	examples

1.	\bsaka\b .* \bhaa	string of 3 words:
		(1) *saka*
		(2) any word, and
		(3) the word *haa* by itself or with suffixes
			saka antee haa;
			saka abana haari;
			saka kabuu haana

2.	saka .* \bhaa\w+	string of 3 words: (1) *saka* (2) any word, and (3) a words beginning with *haa*, but not *haa* by itself	*saka abana haari;* *saka kabuu haana*
3.	(\bsaka\b|\bsa\b) \bpaku\b	all 2 word strings that consist of *saka* or *sa* and *paku*	*saka paku, sa paku*
4.	(\bsaka\b|\bsa\b) .* \bvaha\b	all 3 word strings with (1) *saka* or *sa,* (2) any word (3) *vaha*	*saka tii vaha* *sa tapaku vaha*
5.	(\bsaka\b|\bsa\b) (....)\1 \bhaa	all 3 word strings with (1) *saka* or *sa,* (2) a word with four reduplicated letters (3) the word *haa* or a word beginning with *haa*	*sa natanata haa,* *saka natanata haana*

Table 3: Combinations of the symbols \b, .*. \w+ and (x|y)

Comments on Table 3:

saka/sa ... haa is a discontinuous negation. The last component *haa* can have a suffix that indicates imperfective aspect and person, e.g. *haana, haari, haara.* The formulars above provide data for the following questions:

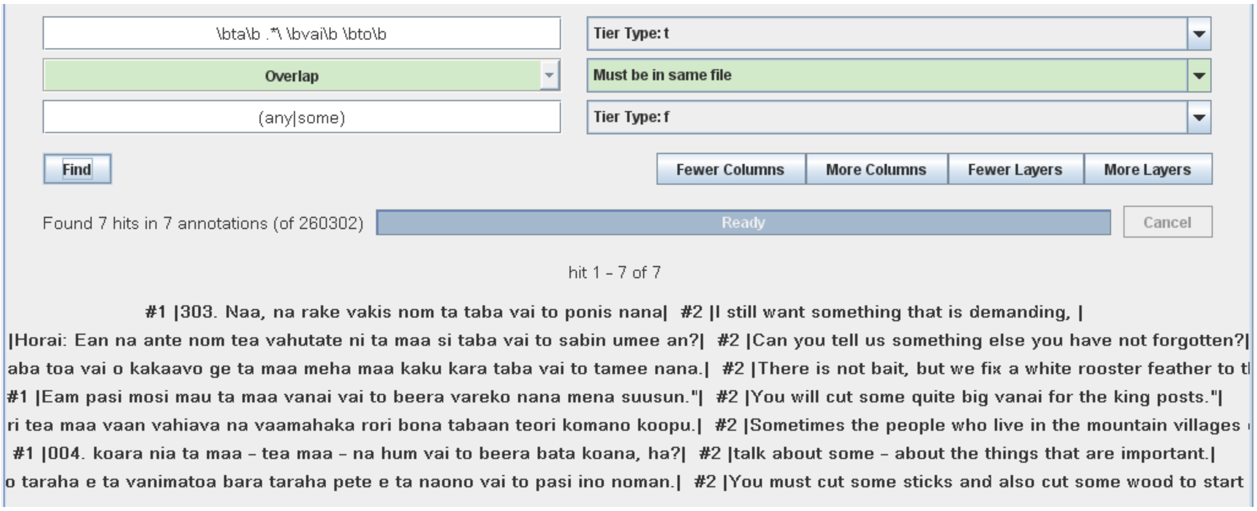
1. Which words are used inbetween *saka* and *haa/haana/haari/haara* ?
2. Which words are used inbetween *saka* and *haana/haari/haara* ?
3. Are there examples for *saka/sa* followed by *paku* ‘do’?
4. Which words are used between *saka/sa* and *vaha* ‘back, also, again, anymore’?
5. Does *saka/sa ... haa* combine with reduplicated words?

4. Multilayer search with regular expressions

Multilayer search is useful if you want to find examples of a homonymous lexical item or functional word as, for instance, the Teop non-specific article *ta* ‘any, some’ which is homonymous with the noun *ta* ‘part’ and the complementizer *ta.*When I came across a sentence in which this non-specific article was followed by the demonstrative pronoun *vai* ‘this’ and a relative clause introduced by *to,* I searched for all examples of this extraordinary construction

(1) *ta* .... X *vai* *to*
ART X DEM REL
‘any/some ... X that

in the corpus using the formula \bta\b .*\ \bvai\b \bto\b on the transcription tier and (any|some) on the translation tier:



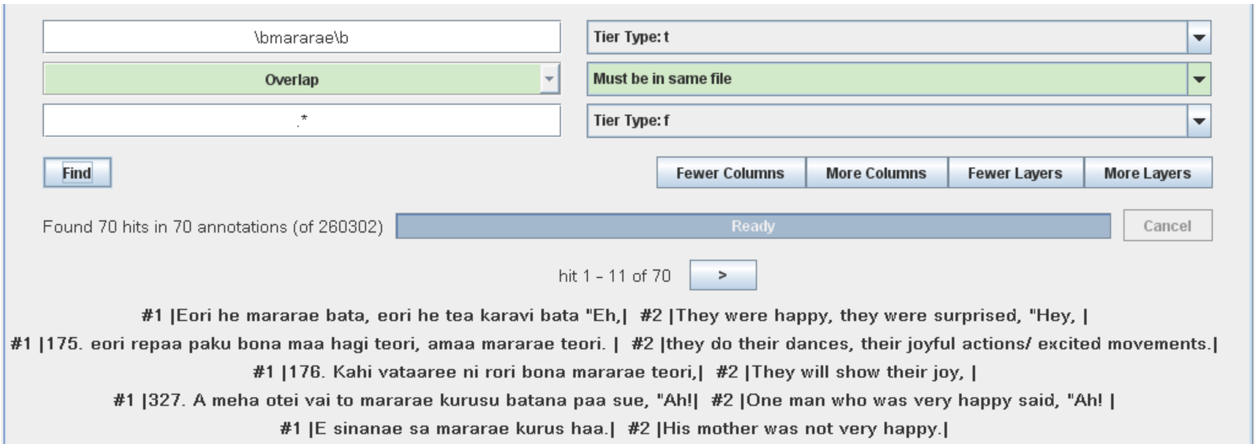
\bta\b .*\ \bvai\b \bto\b — Tier Type: t
Overlap — Must be in same file
(any|some) — Tier Type: f

```
Found 7 hits in 7 annotations (of 260302)

hit 1 - 7 of 7

#1 |303. Naa, na rake vakis nom ta taba vai to ponis nana|  #2 |I still want something that is demanding, |
|Horai: Ean na ante nom tea vahutate ni ta maa si taba vai to sabin umee an?|  #2 |Can you tell us something else you have not forgotten?|
aba toa vai o kakaavo ge ta maa meha maa kaku kara taba vai to tamee nana.|  #2 |There is not bait, but we fix a white rooster feather to t
#1 |Eam pasi mosi mau ta maa vanai vai to beera vareko nana mena suusun."|  #2 |You will cut some quite big vanai for the king posts."|
ri tea maa vaan vahiava na vaamahaka rori bona tabaan teori komano koopu.|  #2 |Sometimes the people who live in the mountain villages
#1 |004. koara nia ta maa – tea maa – na hum vai to beera bata koana, ha?|  #2 |talk about some – about the things that are important.|
o taraha e ta vanimatoa bara taraha pete e ta naono vai to pasi ino noman.|  #2 |You must cut some sticks and also cut some wood to start
```

Figure 1: Multilayer search for *ta* with the translation ‘any’ or ‘some’

This formula means: search within an annotation for all occurences of *ta* meaning ‘any’ or ‘some’ that is first followed by one or more unspecified words and then by the demonstrative *vai* and the relative pronoun *to.*

Multilayer search is also practical, if you do not know the language well and want to search for a word and all its translations. Then you search on the free translation tier with the wild card .* For example, the search for *mararae* gives you the translations ‘happy’, ‘joyful’ and ‘joy’.



\bmararae\b — Tier Type: t
Overlap — Must be in same file
.* — Tier Type: f

```
Found 70 hits in 70 annotations (of 260302)

hit 1 - 11 of 70

#1 |Eori he mararae bata, eori he tea karavi bata "Eh,|  #2 |They were happy, they were surprised, "Hey, |
#1 |175. eori repaa paku bona maa hagi teori, amaa mararae teori. |  #2 |they do their dances, their joyful actions/ excited movements.|
#1 |176. Kahi vataaree ni rori bona mararae teori,|  #2 |They will show their joy, |
#1 |327. A meha otei vai to mararae kurusu batana paa sue, "Ah!|  #2 |One man who was very happy said, "Ah! |
#1 |E sinanae sa mararae kurus haa.|  #2 |His mother was not very happy.|
```

Figure 2: Multilayer search for *mararae* with any translation.