

Humboldt-Universität zu Berlin, Seminar für Afrikawissenschaften
Linguistisches Kolloquium
October 26th, 2010

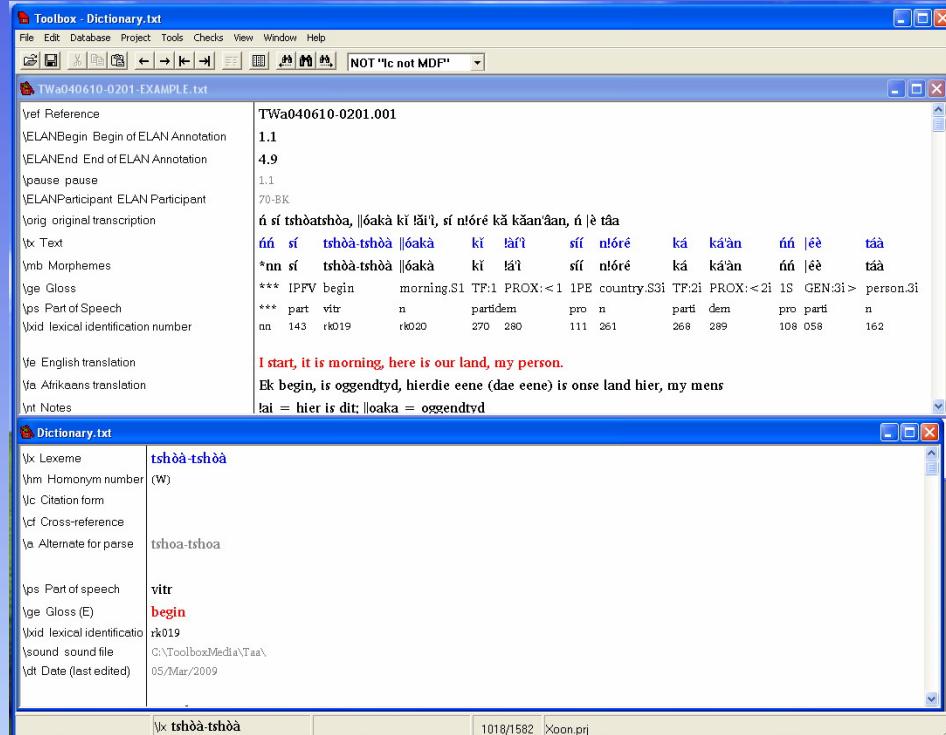
*Two language documentation tools in Khoisan
research:*

2) A multi-dialectal Toolbox project

Christfried Naumann
naumann@eva.mpg.de

Structure

3. A multi-dialectal Toolbox project



D - d	
Dàálá	n. Daala, Daala.
dáhá, dáhásíké	n. jacket, coat, jas. Prdm: 1 / 2ii.
dám kM	vtr. name, call, term, designate, noem.
dáng kM	vprep. chop, cut, kap.
Dáindá	n. PN 3ii. Dáindá sí fáy sábf Danda carries the blanket (on her shoulders). Danda dra die kombers.
dáo kM	vprep. suppose, surmise, dink, sé.
dápná, dápnásiké	n. green plant(s), plant(s) with woody tissue, groen plante. Prdm: 2ii / 2ii.
Dertien	n. Pn 13 (largest settlement in the corridor, SE Omahcke). Pos Dertien. Prdm: 1.
dáhábé	vtr. be soft (e.g. meat), sag wees. Njep! Cá si dhábé. The meat of the springbok is soft. Die springbok se vleis is sag.
dáhábí-dhábí, dhábí-dhábiké	n. butterfly, skoenlapper. Dhábí-dhábiké si dhábí. The butterflies are flying. Die skoenlappers vlieg. Prdm: 1 / 2ii?
dóla, dólárké	n. dollar S1, dollar P2ii. dollar. Prdm: 1 / 2ii.
dómá n M	vitr. lurk, lay in ambush for, spy out. loer vir.
dóqhrM	vtr. to wash, clean (animate object or body part, e.g. a child, teeth), was, skoommaak (n mens, liggaam). dé si dhóqré Ogáé She washes the child. Sy was die kind. dé si dhóqré tåa She washes people. Sy was mense.
dóxter	n. physician S2 dokter.
dómá	vitr. smile, glimlag. tåa káké si dómá This person is smiling. Daai mens glimlag.
dram	adv. therefore, daarom.
dúa, dúašké	fruit S1, fruit P2ii. vrug(te). Prdm: 1 / 2ii.
dxám	vtr. smear, smeer.
dxaré	vitr. slip, gly.
dxóré	See: dxóré. n. guts, intestines. derms. Prdm: 2ii.
dxúf, dxóré	n. gut; large intestine. derm; die grote darm. Prdm: 1 / 2ii.
dzáhó, dzálré	n. borehole, borgat. Prdm: 2ii / 2ii.
dzáqmí	vitr. be coloured, bout wees. Måri ká dzáqmí kán, áaf si gláhr líhm Die coloured goat is climbing the dune. Die bok wat bout is klm die
E - e	
-e	demsf. <3i, <3ii.
-é	vsfpro. 3i.
2	Taa English Afrikaans Dictionary

3. A multi-dialectal Toolbox project

- Two objectives:

A) Dialect-specific morpho-syntactic analysis (interlinearisation)

The screenshot shows two windows from a software application. The top window is titled 'Toolbox - Dictionary.txt' and contains a table of ELAN annotation tags and their values. The bottom window is titled 'Dictionary.txt' and contains a table of lexical entries. An arrow points from the word 'qáí' in the dictionary window to its corresponding entry in the toolbox window.

Toolbox - Dictionary.txt

\ref Reference	TWa040610-0201.004
\ELANBegin Begin of ELAN Annotation	11.6
\ELANEnd End of ELAN Annotation	14.3
\ELANParticipant ELAN Participant	70-BK
\tx Text	túù qáí nlúmí kê þì !Úm þáññ áàn ká láññ
\mb Morphemes	túù
\ge Gloss	people.P4
\ps Part of Speech	n
\xid Lexical identification number	165
\te English translation	It is where the people had settled along the streaks of dunes,
\fa Afrikaans translation	is waar die mense gebly het die duine se streepe

Dictionary.txt

\x Lexeme	qáí
\a Alternate for parse	qai
\hm Homonym number	(W)
\c Citation form	
\cf Cross-reference	
\ps Part of speech	adv
\ge Gloss (E)	PST3
\xid Lexical identification	rk030
\sound sound file	
\dt Date (last edited)	27/Aug/2008
\de Definition (E)	long time ago, in ancient times (e.g. 10 years ago or longer)
\ue Usage (E)	
\dv Definition (v)	

3

3. A multi-dialectal Toolbox project

- Two objectives:

A) Dialect-specific morpho-syntactic analysis (interlinearisation)

The screenshot shows the 'Toolbox' application interface with two main windows:

- Top Window (Toolbox - TWa040610-0201-EXAMPLE.txt):** This window displays a text file with morphological analysis and interlinearization. The text is:

\ref Reference TWa040610-0201.004
\ELANBegin Begin of ELAN Annotation 11.6
\ELANEnd End of ELAN Annotation 14.3
\ELANParticipant ELAN Participant 70-BK
\x Text túù qáí nlúm kê ji ÍUm tháin àan ká láan
\mb Morphemes túù qáí nlúm kê ji ÍUm tháin àan ká láan
\ge Gloss people.P4 PST3 live LOC:3ii> GEN:1> GN.1 streak.2i 2i.O TF:2i PROX:<2i
\ps Part of Speech n adv vitr parti parti n n pro parti dem
\xid lexical identification number 165 rk030 rk028 034b 055 rk021 rk029 114 268 308
\fe English translation It is where the people had settled along the streaks of dunes,
\fa Afrikaans translation is waar die mense gebly het die duine se streepe
- Bottom Window (Dictionary.txt):** This window shows a detailed view of a lexeme entry for 'qáí'.

\x Lexeme	qáí
\a Alternate for parse	qai
\hm Homonym number	(W)
\c Citation form	
\cf Cross-reference	
\ps Part of speech	adv
\ge Gloss (E)	PST3
\xid Lexical identification number	rk030
\sound Sound file	
\dt Date (last edited)	27/Aug/2008
\de Definition (E)	long time ago, in ancient times (e.g. 10 years ago or longer)
\ue Usage (E)	
\dv Definition (V)	
\gr Gloss (r)	baie lank gelede
\gn Gloss (n)	

3. A multi-dialectal Toolbox project

- Two objectives:

B) A multidialectal dictionary

Q - q	
qáán <i>n.</i> gum. <i>gom.</i> <i>Prdm:</i> 2i.	<i>hom.</i>
qáé , qárú <i>n.</i> mother S3i. <i>moeder.</i> <i>'fíí qáé</i> my mother <i>my moeder</i> <i>Prdm:</i> 3i / 4.	qáí <i>adj.</i> clever. slim.
qàhàn <i>adv.</i> much. <i>baie.</i> <i>éé qàhàn sí qx'áí</i> She is laughing a lot. <i>Sy lag baie.</i>	qháà-+"ómàkê <i>See:</i> qháà-+"úì . <i>n.cp1.</i> caracal P2ii. qháà-+"úì , qháà-+"ómàkê <i>n.cp1.</i> caracal. <i>rooikat.</i> <i>Prdm:</i> 1 / 2ii.
qàhrM <i>vtr.</i> dig out. <i>grou,</i> <i>grawe.</i> <i>éé sí qàhrí nñùqùn</i> He is digging (out) the hare. <i>Hy grawe die haas uit.</i>	qhàè <i>interj.</i> oh! (sound of embarrassment). <i>oh!</i>
qái <i>adv.</i> long time ago, in ancient times (e.g. 10 years ago or longer). <i>baie lank gelede.</i>	qhai <i>See:</i> kuri. <i>n.</i> year P?
qáí <i>vitr.</i> fall S. <i>val.</i> <i>qáí túù</i> fall down (sg). <i>val.</i> <i>!óyè qáí túù</i> A stone is falling down. <i>'N klip val.</i>	qhárn <i>adv.</i> really, truly, seriously. <i>rérig,</i> <i>ernstig.</i> <i>Táä</i> <i>qhárn qáí túù.</i> The person has really/seriously fallen (down). <i>Die mens het rérig/ernstig geval.</i>
qái kâ <i>vprep.</i> fall P. <i>val.</i> <i>qái kâ +q'aà</i> fall down (pl). <i>val.</i> <i> árn áan kâ sí qái</i> <i>kâ +q'aà</i> CAMEL THORN TREES are falling down. <i>Dis kameeldorings wat neerval.</i>	qhárn <i>n.</i> truth, true. <i>waarheid,</i> <i>waar.</i> <i>Üú sí nñáre kí</i> <i>qhárn.</i> They tell the truth. <i>Hulle praat die</i> <i>waarheid.</i> <i>Prdm:</i> 1.
qáín <i>vitr.</i> be beautiful S. <i>mooi wees.</i> <i>blðrn qáín</i> The flower is beautiful. <i>Die blom is mooi.</i>	qháré <i>adv.</i> at noon, at midday, in the warm time of the day. <i>in die middag.</i> <i>Túù á qháré sáí.</i> The people arrived at noon (in the warm time of the day). <i>Die mense het in die middag gekom.</i>
qáí-sààn-qáí <i>n.cp1.</i> killer. <i>doodmaker.</i>	qhàùkâ <i>adv.</i> almost, nearly. <i>amper,</i> <i>byna.</i> <i>Táä</i> <i>qhàùkâ qáí túù.</i> The person has nearly/almost fallen (down). <i>Die mens het amper geval.</i>
qàm <i>vitr.</i> suck. <i>suip.</i>	qoa <i>part.</i> long ago, in the past (e.g. several months or some years). <i>lank gelede.</i>
qányá <i>vitr.</i> to be beautiful. <i>baie mooi wees.</i> <i>Èè á</i> <i> hðánsíkéè sí qányá.</i> Her hair (pl) is beautiful. <i>Haar hare is baie mooi.</i>	qòqà <i>adv.</i> long ago, in the past (e.g. several months or some years). <i>lank gelede.</i> <i>kùhmí qòqá g húm</i> The cow was pregnant. <i>Die beest was dragtig.</i>
qàqàn <i>adv.</i> still. <i>nog.</i> <i>éé qàqàn sí áan</i> He is still eating. <i>Hy eet nog.</i>	qárá <i>adj.</i> young, little, small. <i>junk klein</i>

3. A multi-dialectal Toolbox project

- How to meet these two objectives?

The screenshot illustrates a multi-dialectal Toolbox project with three windows:

- Main Window (TWb080410-0101.txt):** Shows a transcription table with columns for reference, ELAN time, duration, pause, original text, and various linguistic features like verb tense and part of speech. A red circle highlights the word "n|àí". Below the table is a gloss in English and Afrikaans.
- Dictionary Window 1 (Dictionary.txt:1):** Shows entries for the lexeme "n|áM". The entry includes citation forms "n|áM" and "n|á", parts of speech "vtr", and a gloss "see". A red box contains the text: "variant of the 'N/oha and East !Xoon varieties'".
- Dictionary Window 2 (Dictionary.txt:2):** Shows entries for the lexeme "n|àM". The entry includes citation forms "n|àM" and "n|à", parts of speech "vtr", and a gloss "see". A red box contains the text: "variant of the West !Xoon variety".

A question mark is placed between the two dictionary windows, indicating a comparison or relationship between the different variants.

3. A multi-dialectal Toolbox project

- One possible solution: Add field \var "variety":

DictPersonal.txt	
\lx Lexeme	qáí
\var variety	(W)
\a Alternate for parse	qai
\lc Citation form	
\cf Cross-reference	
\ps Part of speech	adv
\ge Gloss (E)	PST3
\lid Lexical identification number	rk030

⇒ not successful (problems with MDF export)

3. A multi-dialectal Toolbox project

- Our suggestion: Use the marker \hm "homonym number":

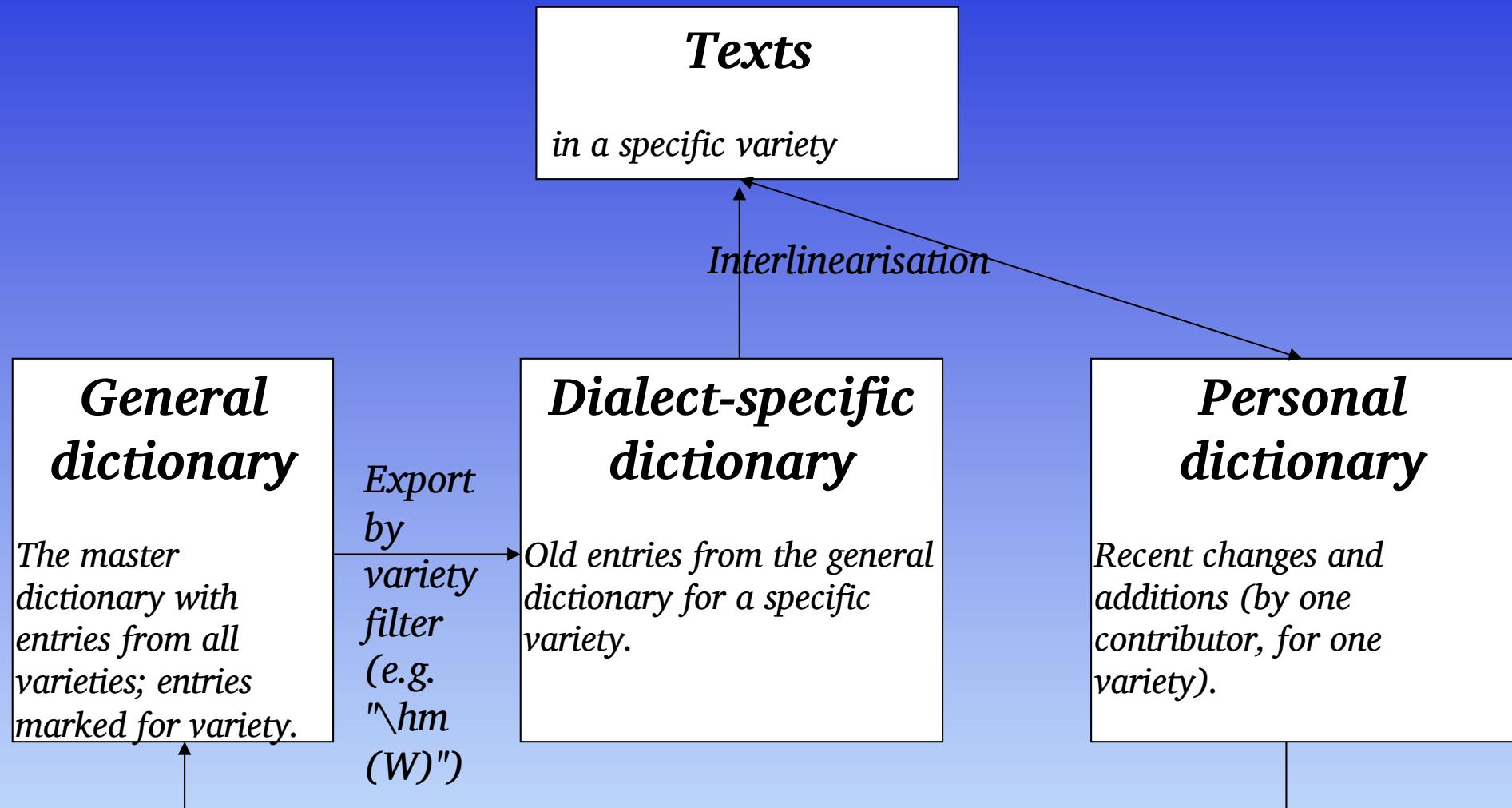
\lx Lexeme	xáè	Morpheme (tones marked)
\hm Homonym number	(W)	Dialect
\lc Citation form	sí xáè	Citation form (with affixes, tones)
\pl Plural form	sí xààn	Plural form (with affixes, tones)
\cf Cross-reference		(References to related entries)
\a Alternate for parse	xae	Alternate for parse – toneless
\a Alternate for parse	sí xáè	
\u Underlying for parse	sí- xáè	
\ps Part of speech	n	part of speech
\dv Definition (v)	3i/2ii	noun class of singular/plural
\pdv Paradigm form	-è/-àñ	
\ge Gloss (E)	man.S3i	gloss with noun class (for interlinearisation)
\lid Lexical identification number	152	lexical ID (for repeated interlinearisations)
\sound sound file	C:\ToolboxMedia\Taa\152.wav	link to sound file
\dt Date (last edited)	15/Jun/2010	
\de Definition (E)	man, male person; husband	English gloss for MDF dictionary
\re Reversal (E)	man ; male ; husband	English gloss for List English – Taa
\ue Usage (E)		
\gr Gloss (r)	man ; eggenoot	Afrikaans gloss for MDF dictionary
\gn Gloss (n)	moíñā	Setswana gloss for MDF dictionary

3. A multi-dialectal Toolbox project

Dictionary.txt	
\ue Usage (E)	
\gr Gloss (r)	man ; eggenoot
\gn Gloss (n)	moínā
\gk gloss Khoekhoe	khòèb, áòb
\sd Semantic domain	02.210
	<i>Khoekhoe gloss for MDF dictionary</i>
	<i>Semantic domain (IDS list)</i>
\nq Notes (questions)	correct?
\nt Notes (general)	cf. TWb080410-0101.032
\ph Phonetic form	[sí χáè]
\va Variant form(s)	
\np Notes (phonology)	
\ng Notes (grammar)	
\na Notes (anthropology)	
\et Etymology (proto form)	
	<i>(notes grammar)</i>
\xv Example (v)	Sí xáè sí sáì.
\xe Example free trans. (E)	The man is coming.
\xr Example free trans. (r)	Die man kom.
\xn Example free trans. (n)	
\rf Reference	TWb080410-0101.032
	<i>example for MDF dictionary</i>
	<i>English translation for MDF dictionary</i>
	<i>Afrikaans translation for MDF dictionary</i>
	<i>reference of example</i>

3. A multi-dialectal Toolbox project

- Taa toolbox projects:



Editor integrates changes (marked for variety) into general dictionary.

3. A multi-dialectal Toolbox project

- Example (interlinearisation)

The screenshot displays four windows from a software application for linguistic analysis:

- TWb080410-0101.txt**: Shows interlinearized text with morphological annotations. The text includes words like 'ní sí', 'n|àí', 'lxùbí', 'kí', 't"áí', 'lqhàà', and 'kìì'. Annotations include \ref, \ELAN, \pause, \orig, \lx, \mb, \ge, \ps, and \xid.
- DictPersonal.txt**: A dictionary window showing lexemes and their forms. It includes entries for 'lx Lexeme' (lxùbí), 'hm Homonym number' ((W)), 'lc Citation form' (lxùbí), 'pl Plural form' (lxubikê), 'cf Cross-reference', 'a Alternate for parse' (lxubi), 'ps Part of speech' (n), 'dv Definition (v)' (1/2ii), 'pdv Paradigm form' (-i), 'ge Gloss (E)' (pool.S1), and 'ge Gloss (E)' (pool.P2ii).
- Dictionary.txt**: A large table of words with various properties. Columns include: \lx Lexeme (e.g., n|àhM, n|àhmà, n|áî, n|àì, n|àín, n|àM, etc.); \hm Homonym number (e.g., (W), (W), (W), (W), (W), (N, ZUT)); \ps Part of: (e.g., vtr, n, parti, parti, ?, vtr); \ge Gloss (E) (e.g., say, game.2i, DAT:<1P, DAT:1>, ?, see); and \xid (lexical identification number, e.g., 851).
- DictDialect.txt**: A smaller table showing dialectal variants. It includes columns: \lx, \hm, \ps, and \ge. Entries correspond to the rows in the Dictionary.txt table, such as n|àhmà (W), vtr, say, game.2i; n|áî (W), n, game.2i, DAT:<1P; and so on.

3. A multi-dialectal Toolbox project

- Example (exported MDF dictionary)

Q - q

qáán (W) *n.* gum. *gom.*

qáé (W, ZUT) Pl: qárú. *n.* mother S3i. *moeder.* **ńń qáé** (W) my mother

qàhàn (W) *adv.* much. *baie.* **èé qàhàn sí qx'ái** (W) She is laughing a lot.

!qàhè (W, ZUT) *vitr.* hunt. *jag.*

qàhrM (W) *vtr.* dig out. *grou, grawe.* **èé sí qàhrí n!ùqùn** (W) He is digging (out) the hare.

qáí (W) *adv.* long time ago, in ancient times (e.g. 10 years ago or longer). *baie lank gelede.*

qáì (W) *vitr.* fall *S. val.*

qáì túù fall down (sg). *val.* **!'óyè qáì túù** (W) A stone is falling down.

qáí kâ (W) *vprep.* fall *P. val.*

qáí kâ †q'aà fall down (pl). *val.* **llám' áàn ká sí qáí kâ †q'aà** (W) CAMEL THORN TREES are falling down.

qáín (N, W) *vitr.* be nice, beautiful. *mooi wees.* **blòm' qáín** (W) The flower is beautiful.

qárà (W) *vitr.* dig. *grou.*

qárM (W) *vtr.* dig, dig out. *grou, graaf, grawe.* **èé sí qánà g†qx'um** (W) He is digging the sand. **Filà sí qárè nō áqè.** (W) Fila is digging out the wild cucumber.

qárú (W) See: qáé. *n.* mothers, parents. *moeders.*

qáyM (W) *vtr.* kill. *doodmaak.* **nlàhng "qáyè" kè-éè** (W) Say "Kill him (3ii)" to him.

qá'i (W) *adj.* clever. *slim.*

qháà-‡'ómàké (W) See: qháà-‡'úì. *n.* *cpI.* caracal P2ii.

qháà-‡'úì (W) Pl: qháà-‡'ómàké. *n.* *cpI.* caracal. *rooikat.*

qhàè (W) *interj.* oh! (sound of embarrassment). *oh!*

qhai (W) See: kuri. *n.* year P?

qhám' (W) *adv.* really, truly, seriously. *rérig, ernstig.* **Táá qhám' qáì túù.** (W) The person has really/seriously fallen (down).

qhám' (W) *n.* truth, true. *waarheid, waar.* **Ùú sí nllárè kí qhám.** (W) They tell the truth.

Acknowledgements

- Alena Witzlack-Makarevich
- Anja Bonitz
- Gertrud Boden
- **Hans-Jörg Bibiko** (*technical supervision, Tsammalex*)
- **Jan Klom** (*original set-up, Tsammalex*)
- Juliane Böttger
- **Kathrin Heiden** (*data East !Xoon + Setswana, Tsammalex*)
- Martina Ernszt
- Nadine Borchardt
- **Peter Fröhlich** (*administration Tsammalex*)
- **Roland Kießling** (*Toolbox project*)
- **Stefan Koch** (*programming Tsammalex*)
- Sven Hubold
- Sven Siegmund
- **Tom Güldemann** (*Toolbox project*)