
**Documenting grammatical tone
using Toolbox: an evaluation of
Buseman's interlinearisation technique**

Stuart McGill

Proceedings of Conference on
Language Documentation & Linguistic Theory

Edited by Peter K. Austin, Oliver Bond & David Nathan

7-8 December 2007 School of Oriental and African Studies, University of London

School of Oriental and African Studies
Thornhaugh Street, Russell Square
London WC1H 0XG
United Kingdom

Department of Linguistics:
<http://www.soas.ac.uk/academics/departments/linguistics>

Hans Rausing Endangered Languages Project:
<http://www.hrelp.org>
elap@soas.ac.uk

© 2007 Stuart McGill

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, on any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the author(s) of that part of the publication, except as permitted by UK copyright law.

ISBN: 978-0-7286-0382-0

This paper can be cited as:

Stuart McGill. 2007. Documenting grammatical tone using Toolbox: an evaluation of Buseman's interlinearisation technique. In Peter K. Austin, Oliver Bond & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory*. London: SOAS. pp. 177-186.

or:

Stuart McGill. 2007. Documenting grammatical tone using Toolbox: an evaluation of Buseman's interlinearisation technique. In Peter K. Austin, Oliver Bond & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory*. London: SOAS. www.hrelp.org/eprints/ldlt_21.pdf

Documenting grammatical tone using Toolbox: an evaluation of Buseman's interlinearisation technique¹

STUART MCGILL

School of Oriental and African Studies

1. INTRODUCTION

For tone languages, particularly those found in Africa, it is often the case that important grammatical distinctions depend solely on the occurrence of different tone patterns on a particular noun or verb. The goal of this paper is to consider how the grammatical contribution of such tone patterns should be represented in the annotations accompanying documentary corpora. After introducing certain tone phenomena from Cicipu (§2), a Benue-Congo language spoken in northern Nigeria, I will challenge the view that it is not appropriate to make explicit the effects of tone patterns in interlinear annotations (§3), and review a technique devised by Alan Buseman of SIL for handling tone using the software program Toolbox² (§4). Based on the evaluation (§5) of this technique as applied to a corpus of Cicipu texts³, I suggest some recommendations for the future development of interlinearisation software (§6).

2. GRAMMATICAL TONE IN CICIPU

GRAMMATICAL tone refers to a tonal change which signals not a lexical difference of the kind shown in (1), but a grammatical difference as in (2).

- | | | | |
|-----|--------------------|---------------|------------------|
| (1) | (a) <i>káayá</i> | 'room' | H H ⁴ |
| | (b) <i>káayà</i> | 'bean' | H L |
| (2) | (a) <i>índúkwà</i> | 'I went' | L H L |
| | (b) <i>índùkwà</i> | 'I should go' | H L L |

¹ Special thanks are due to Alan Buseman for posting details of his Toolbox technique in response to an initial query by Sophie Salfner and myself, and to Oliver Bond, Tucker Childs, and Alan Buseman again for helpful comments on an earlier version of this paper. I would also like to thank Heidi Rosendall for help with some technical details of Toolbox.

² <http://www.sil.org/computing/toolbox>

³ This material was collected during a six-month field trip by the author from September 2006 until March 2007, funded by a Field Trip Grant from the Hans Rausing Endangered Languages Documentation Project. Unless otherwise stated, all examples in this paper are from Cicipu.

⁴ Cicipu has two contrasting levels of tone (written H and L here) plus a falling tone HL, which can usually be analysed as a sequence of H plus L. Other abbreviations used in this paper are 2 = second-person, 3 = third-person, ACC = accusative, AG = agreement, AOR = aorist, FUT = future, GEN = genitive, HAB = habitual, IRR = irrealis, IMP = imperative, INJ = injunctive, NC = noun class, NEG = negative, PL = plural, RLS = realis, SG = singular.

The tone patterns on the nouns in (1) make no isolatable contribution to the meaning of the word. Consequently the tones on the lexemes *káayá* and *káayà* are included in the lexicon along with the segmental phonemes, and the corresponding word-forms are treated as single morphemes during interlinearisation. In contrast, the examples in (2) share a common component of meaning, namely a GOING event involving the speaker as theme. The tone patterns do contribute a separable component of meaning, either REALIS in (2a) or IRREALIS in (2b). It will be argued in §3 that the contribution of tone patterns such as these to the meaning of the utterance should be made explicit in interlinear annotations.

While they are conceptually distinct, lexical and grammatical tone can interact with each other when a tone pattern associated with a grammatical category is superimposed on a noun or verb with inherent lexical tone. Cicipu nouns provide a good example. In isolation or when functioning as subject, each noun is pronounced with a particular tone pattern, and these have to be learned individually. However they are also affected by grammatical tone, in that they may take on different tone patterns depending on the syntactic context. Verbs on the other hand are inherently toneless, and the tones with which they surface are determined entirely by grammatical properties such as mood or aspect.

This interaction between lexical and grammatical tone can be a problem for interlinearisation, as will be apparent in the discussion of Buseman's method in §4-5. Before turning to the more general discussion, however, it will be helpful to consider some examples of grammatical tone in more detail:

- (3) (a) *ù-hyâan* *ù-dúkwà*
 3SG-say\RLS 3SG-go\RLS
 'he said he went'
- (b) *ù-hyâan* *ú-dùkwà*
 3SG-say\RLS 3SG-go\IRR
 'he said he should go'
- (c) *ùhyâan* *Ø-dúkwà*
 3SG-say\RLS 2SG-go\RLS
 'he said you went'
- (d) *ùhyâan* *dùkwá*
 3SG-say\RLS go\IMP
 'he said "go!"'
- (4) (a) *à-câa* *cì-kùlù* *tì-wómó*
 3PL-give\RLS NC6-tortoise\ACC NC6-chief
 'they gave the chieftancy to the tortoise'

- (b) *â-câa* *cì-kúlú* *tí-wòmò*
 3PL-give\RLS NC6-tortoise AG6-chief\GEN
 'they gave the tortoise of the chief'

These examples illustrate three different grammatical alternations involving just tonal changes. In each case, the segmental material remains the same, but the different tone patterns superimposed on (or interacting with) the words give rise to different grammatical categories, and hence different meanings. In (3) the verb can be found in three different moods, depending on the tone pattern. In (3a) the verb in the subordinate clause *dukwa* has a H L pattern with a low-tone third-person prefix, which together indicate realis mood, while in (3b) *dukwa* has a L L pattern with a high-tone prefix, indicating irrealis mood. Examples (3c) and (3d) illustrate a similar contrast, this time between realis and imperative, and involving the second-person. Example (4) demonstrates two further tonal processes. First, note that the tone on *ci-kulu* 'tortoise' is different in (4a) and (4b). In (4b) the citation form H H⁵ appears, but in (4a) the tone is L L, a pattern which occurs with certain nouns in the complement position of the verb phrase. Secondly, the morpheme *wòmò* 'chief' also varies in tone. This time (4a) shows the citation pattern of the class 6 noun *tì-wòmò* 'chieftancy' (H H), while (4b) shows a perturbed form (L L) in the genitive.

3. INTERLINEAR FORMATS FOR GRAMMATICAL TONE

Most documentation projects will include interlinearised annotations. According to Schultze-Berndt, 'in an annotated corpus, it is also recommended practice to include interlinear glosses for all or at least part of the transcriptions' (2006: 239). Others recommend that only a small amount of the transcribed text should be interlinearised, in which case it is especially important that the output is well-structured. The question to be addressed here is how the kind of tone patterns illustrated in §2 should be represented in such interlinear texts. An obvious starting point for the discussion is the Leipzig Glossing Rules (Bickel, Comrie, and Haspelmath 2004), a set of widely-accepted conventions for interlinear glossing. The relevant rule here is 4D (2004: 4) which states that 'if a grammatical property in the object-language is signaled by a morphophonological change of the stem (ablaut, mutation, etc.), the backslash is used to separate the category label and the stem gloss'. This rule was exemplified in examples (3) and (4).

Lehmann (2004b) includes the same rule (R20), which 'distinguishes [stem-change processes] from other morphological processes, but not from each other. Such a morpheme can hardly be signaled in the L1 representation' (2004a: 26 [2004b]) [L1 refers to the object language, and L2 the metalanguage]. As a result of this indeterminacy, in the annotations presented in (3) it is not apparent what it is in the text that contributes the verb mood. The process signalled by the

⁵ The tone on prefixes is almost always predictable. H H here refers to the tones on the noun root.

backslash could refer to the first tone, the second tone, the tone pattern as a whole, or even a vowel change or consonant mutation.

The backslash notation treats a form such as *dúkwa* ('go\RLS') as formally unsegmentable. Nevertheless the two units *dukwa* and H L can be conceptually isolated and directly linked to separate components of meaning. Without making this link explicit, the contributions of the tone patterns in examples (3) and (4) can only be interpreted when viewed as members of a paradigm of word-forms, which an interlinear annotation does not provide. Tone-marked transcriptions are better than no tone-marking at all, and are of obvious benefit to linguists familiar with the language. However this practice by itself does not always meet Lehmann's criterion that the annotation should 'represent the language for those who do not have access to the language itself' (2001: 88).

Lieb and Drude (2000) have criticised traditional interlinear glossing in that it does not allow the degree of annotation necessary for best-practice language documentation. They provide an alternative annotation technique called 'Advanced Glossing' (AG) which provides numerous tiers for the representation of different kinds of linguistic data. Of relevance here is the fact that the segmental and suprasegmental parts of a word can be separated in the representation, and different glosses can be applied to each. Although AG allows the link between tone pattern and gloss to be made explicit, it does so at the expense of readability and conciseness, as well as being time-consuming to implement (Drude 2003, Schultze-Berndt 2006: 251).

It is possible to imagine further complex annotation systems which would bring out the contribution of tone patterns, in particular using insights from autosegmental phonology (Goldsmith 1990), for example by splitting either the transcription and gloss into two separate lines, one for the segmental tier and one for the tonal tier. However such systems would likely suffer from similar problems as AG regarding their readability and the time taken to produce them. Also the use of autosegmental formalism would introduce a theoretical sophistication into the annotations, something Lehmann (2001) argues against.

The methods discussed so far suffer either from a lack of explicitness or from undesirable complexity. The format produced by Toolbox provides an opportunity for compromise, crucially because it uses one more tier than standard IMGs. It is worth making the differences between the two formats explicit. Lehmann (2004b) observes that standard IMGs consist of three lines, as in examples (3) and (4). The first line is a morphophonemic (or orthographic) transcription in L1, with morph breaks indicated by hyphens. The second is a morphemic representation, where the L1 morphemes are given mnemonic names in the L2 metalanguage. The third line is the L2 free translation. By contrast, Toolbox interlinear glosses have (at least) four lines:

(5) \tx⁶ ùdúkwà
 \mb ù- dúkwà
 \ge 3SG- go\RLS
 \ft ‘He went’

The first, third, and fourth correspond to the three lines in a traditional IMG. The second line \mb shares properties with both the \tx and \ge lines. Like the transcription it is written in L1, but like the gloss it is morphemic rather than morphophonemic. This extra line provides an opportunity to link tone patterns and glosses without disturbing the integrity of either the L1 morphophonemic transcription or the L2 morphemic gloss, as in (6):

(6) \tx ùdúkwà
 \mb ù- dukwa- H L
 \ge 3SG- go RLS
 \ft ‘He went’

The resulting annotation is concise, intuitive, and makes clear the link between the H L tone pattern and realis mood. This solution relies on the fact that although the Toolbox \mb line is populated with the *names* of morphemes, the fact that these names are written in L1 serves as a pointer to the form of the corresponding morph represented in the \tx line. Ideally the tone patterns represented in the \mb line should remain in a one-to-one relationship with the gloss. For example, the realis forms of monosyllabic Cicipu verbs have a falling (HL) tone pattern, and those of trisyllabics have H L L. The tone patterns in the \mb line should in all cases be written as H L to avoid mixing morphemic and morphophonemic representations on the same tier. In certain cases the discrepancy between the 'citation' form of the tone pattern and the realisation in a particular example may be too great for the \mb line to serve as a reliable mnemonic, in which case a representation in the form of (6) may end up obscuring rather than elucidating the link between tone and meaning. This problem is taken up again in §6.

Before turning to how such annotations can be produced using Toolbox, it should be admitted that the format in (6) is not compatible with all morphological theories, since it treats the realis tone pattern as a 'suprasegmental morpheme'. Such analyses were common under the Item-and-Arrangement morphological model favoured by American structuralists (Hockett 1954, Matthews 1974: 79), but are less popular today. Nevertheless it is not always easy to see how the insights of more modern theories can be concisely represented in interlinear form. As Lieb and Drude (2000) point out, the interlinear format itself is inherently biased towards the Item-and-Arrangement model.

⁶ This paper uses standard Toolbox field markers: \a = alternate form, \ft = free translation, \ge = gloss (English), \lx = lexeme, \mb = morpheme break, \tx = text.

4. BUSEMAN'S METHOD

The interlinear format set out in (6) (minus field markers) does not presuppose the involvement of Toolbox. It could equally well be used in the production of interlinear texts by other means, and it seems a sensible presentation format to use in descriptive works where the contribution of grammatical tone is at issue. Nevertheless this particular format is included here because it is the one generated using a technique for interlinearising grammatical tone devised by Alan Buseman⁷. Buseman's technique allows annotations such as the one shown in (6) to be generated without sacrificing Toolbox's semi-automatic interlinearisation, a facility that makes the program indispensable to many linguists. The method is briefly summarised here, and then evaluated in §5.

The main difficulty in parsing tone through Toolbox is that although tones can vary independently of the segmental material to which they are attached, Toolbox by default understands them as an intrinsic part of the string of characters: for example *dùkwá* is composed of d + u + ` + k + w + a + ⁸. The challenge is to extract from such a string a morpheme break line of *dukwa* ('go') + L H (IMP). The crucial first step of Buseman's method is to add tone marks to the lexicon as both suffixes and infixes. So for example L will be represented as:

```
(7)  \lx   -L
      \a   `
      \a   -
      \ge  tone.mark
```

This technique relies on the fact that when Toolbox detects an infix, it extracts it out to the end of the word (by default), and so if we add lexical entries such as (7) for both low and high tones then *dùkwá* will be parsed as:

```
(8)  \tx   dùkwá
      \mb   dukwa   -L           -H
      \ge   go      -tone.mark  -tone.mark
      \ft   'go!'
```

The final step is to register the combination of L + H as a named tone pattern, in this case the imperative, by adding a further lexical entry as in (9). The resulting interlinear parse shown in (10) is in precisely the same format as (6).

```
(9)  \lx   -L -H
      \ge   IMP
```

⁷ <http://www.sil.org/computIng/toolbox/extras.htm> (posted in January 2007).

⁸ Using combining diacritics rather than pre-composed characters is essential for this technique.

(10)	\tx	<i>dùkwá</i>	
	\mb	<i>dukwa</i>	-L -H
	\ge	go	IMP
	\ft	‘go!’	

5. EVALUATION

This section provides an evaluation of the technique just described, based on its application to over two hours of transcribed and tone-marked spoken Cicipu texts. Overall the technique has been successful, and I continue to use it for interlinearisation. There were however certain difficulties in applying the technique, as well as more general issues. These will now be described in turn.

5.1. Interaction of lexical and grammatical tone

Recall from §2 that Cicipu nouns have inherent lexical tone, as well as undergoing tonal perturbation in certain grammatical contexts. So for example *táari* ‘(stone)’ becomes *tàari* in the genitive. To allow this to parse as *taari*-GEN it was necessary to add an alternate (\a) toneless form *taari* to the lexicon, a step which could be automated for all nouns if necessary.

5.2. One-infix-per-word restriction

For performance reasons Toolbox contains an inherent restriction that allows only one infix per word to be parsed. Cicipu verbs ending with a digraph in the transcription therefore failed to parse, for example diphthongs, or long or nasal vowels e.g. *ù-tínàà* ‘(he swore)’ and *ù-wónsòn* ‘(it barked)’. As a workaround, the last accent in such words can be moved to the end of the word, as in *ù-tínaà*. Fortunately Cicipu roots only have CV structure, but for languages with CVC syllables it may look ungainly to place tone marks over codal consonants.

Trisyllabic roots initially failed to parse for the same reason as forms ending with a digraph. This is because the first two tones fall in the middle of the string of characters, but only one of them can be handled as an infix. This time the problem cannot be solved by shifting accents, but requires alternate forms (\a) to be added to the lexical entries of all trisyllabic roots, with the middle vowel marked for tone (e.g. *jungònu* ‘shut’). The form *júngònù* will then be parsed as *jungonu*-RLS, with the first and third tones being glossed as the realis tone pattern H L, the second tone being already included in the alternate form. Adding the extra forms was less arduous than it might have been since very few Cicipu roots have more than two syllables, and there are only two contrastive tones to consider. For languages with more complex word-structure or tone systems, many more entries may be required.

Languages with ‘real’ infixes are especially problematic given the one-infix-per-word restriction. Cicipu has both iterative and causative verbal infixes, and verb-forms containing these could not be parsed without removing either the tone pattern or the infix from the transcription. Buseman proposes a solution to problems caused by the infix restriction in what he refers to as his ‘second

method', although it actually works in conjunction with his first. This solution involves applying a batch program (CCT) prior to interlinearisation to extract the tone marks from the text and place them at the end of each word. So for example *júngònù* would become *jungonú` `*, which would then be parsed as *jungonu*-RLS without the need to treat any of the tone marks as infixes. The problem with this method is that the batch program has no way of determining parts of speech. Therefore the process is applied indiscriminately to all words of all lexical categories, which is unlikely to be appropriate – unless, of course, only grammatical tone is marked in the transcription and not lexical tone. This could well be the case for an orthographic transcription (see Bird 1999 for orthographies of Kako and Etung which mark only grammatical tone), in which case this technique will be helpful. However it will not work with the morphophonemic transcriptions which are generally recommended for language documentation.

5.3. Spreading

I have not mentioned tone spreading in this paper since in Cicipu it has not been problematic with respect to interlinearisation. What spreading there is seems to be local and mainly affects affixes rather than roots. This can be handled straightforwardly (if unrevealingly) by adding alternate forms to the closed class of affixes. For languages which allow spreading across morpheme or word boundaries, it is hard to see how this could be handled in Toolbox.

5.4. Complexity

At this stage of the Cicipu documentation project the interlinearisation set-up is stable and, other than the conventions for transcribing word-final digraphs, transparent to the user. Nonetheless Buseman's basic technique together with the workarounds I have found necessary to apply amount to a significant increase in the complexity of the interlinearisation process. This is not necessarily a problem for long-term language documentation – as Alan Buseman (p.c.) has pointed out, the final clarity of the interlinearised text is more important than how it was arrived at, since the hypothetical linguist of five hundred years time will not be using Toolbox⁹. It should nevertheless be made clear that the extra \a fields in the lexicon were put there for technical rather than linguistic reasons. One way to do this would be to use a separate marker altogether for such alternate forms (e.g. \at), and to add this to the 'Markers to find' list for parsing.

6. SOFTWARE RECOMMENDATIONS

Some of the problems mentioned in §5 can be overcome or mitigated by further adjustments within Toolbox, but others are more serious and would require a change to the program, or the development of new software. Given the current

⁹ It may be more of a problem in the short-term if the solution proposed here has to be migrated to another program. FieldWorks Language Explorer is designed to be less flexible than Toolbox and does not support the technique outlined here (Heidi Rosendall, p.c. 2007).

uncertainty about the future of Toolbox, there seems little point in making suggestions specific to that program. Nevertheless there are general recommendations that can be made to the designers of future interlinearisation software.

First is a plea that developers consider the autosegmental model of phonology from the outset of program design. In particular, the option to specify a subset of characters (e.g. ´ ˘ ˘) to be treated independently of the remainder is highly desirable, so that tone marks could participate as a *separate* input to the parsing process (as was done in SIL's TonePars program, see Black 1997). Thus word-forms such as *ùdúkwa* could contribute two separate inputs to the parser: *udukwa* together with ´ ˘ ˘, as well as the unanalysed input *ùdúkwa*. This would allow the automatic parsing of verb forms in Cicipu and similar phenomena involving purely grammatical tone. Handling Cicipu nouns would also be simpler, because once the distinction has been made between the segmental and suprasegmental representations of word-forms and lexemes, the ability to specify rules which handle the interaction of lexical and grammatical tone reduces to the familiar Toolbox technique of setting up underlying/alternate forms. Naturally these rules could be constrained according to the principles of autosegmental phonology.

The second recommendation concerns the format of the interlinearised text. As was observed in §3, under certain conditions the L1 'intermediate' tier (∅mb in Toolbox) may end up carrying an unfortunate mix of two different kinds of representations: the usual morphemes plus the tone patterns, which are actually more like suprasegmental 'morphs'. One way to avoid this *mélange*, but retain the explicit link between tone pattern and meaning, would be to allow the option for this intermediate tier to be *morphophonemic* and hence populated with allomorphs rather than morphemes. Even in the case of allomorphs which are quite distinct from the citation form of the morpheme, the link to the morpheme would still be recoverable from the mnemonic properties of the metalanguage gloss, and of course this is exactly how standard (i.e. non-Toolbox) IMG functions. Although the outputs from the Toolbox parsing and glossing processes are configurable, certain technical difficulties mean that the program is unable to produce such a tier. However in general this functionality should be straightforward to implement.

7. SUMMARY

This paper has considered several options for the interlinearisation of utterances involving grammatical tone. The guiding assumption has been the view that the annotation accompanying a documentary corpus should 'represent the language for those who do not have access to the language itself' (Lehmann 2001: 88), and it was argued that strict adherence to the conventions in Lehmann (2004a [2004b]) and Bickel, Comrie, and Haspelmath (2004) can obscure the linguistic structure of constructions involving grammatical tone. There is a tension between the readability of annotations and their feasibility with respect to timescales on the

one hand, and the degree of linguistic structure represented therein on the other. The technique devised by Alan Buseman is suggested as compromise between more complex techniques such as Advanced Glossing which make detailed linguistic annotation possible at the expense of visual compactness and speed of processing, and simpler techniques which fail to capture the link between tone patterns and meanings at all. Although as pointed out in §5 there are deficiencies with the method, it is nevertheless proving valuable for the continued documentation of Cicipu. As well as the recommendations made to software developers in the previous section, I hope that this paper will benefit other linguists working on tone languages by helping them to judge the applicability of Buseman's method to their own documentary corpora.

REFERENCES

- Bickel, Balthasar, Bernard Comrie & Martin Haspelmath. 2004. *The Leipzig glossing rules. Conventions for interlinear morpheme by morpheme glosses*. Leipzig: Max-Planck-Institut für Evolutionäre Anthropologie.
- Bird, Steven. 1999. Strategies for representing tone in African writing systems. *Written Language and Literacy* 2, 1-44.
- Black, H. Andrew. 1997. TonePars: A computational tool for exploring autosegmental tonology. *SIL Electronic Working Papers 1997-007*. <http://www.sil.org/silewp/1997/007/SILEWP1997-007.html>
- Drude, Sebastian. 2003. Digitizing and Annotating Texts and Field Recordings in the Awetí Project. Paper presented at Third EMELD conference on "Digitizing and Annotating Texts and Field Recordings", Lansing, Michigan. Available online: <http://www.emeld.org/workshop/2003/paper-drude.html>.
- Goldsmith, John. 1990. *Autosegmental and metrical phonology*. Oxford: Blackwell.
- Hockett, Charles. 1954. Two models of grammatical description. *Word* 10, 210-34.
- Lehmann, Christian. 2001. Language documentation: a program. In Walter Bisang (ed.), *Aspects of typology and universals*, 83-97. Berlin: Akademie Verlag.
- Lehmann, Christian. 2004a. Interlinear morphemic glossing. http://www.uni-erfurt.de/sprachwissenschaft/personal/lehmann/CL_Publ/IMG.PDF.
- Lehmann, Christian. 2004b. Interlinear morphemic glossing. In Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung. 2. Halbband*, 1834-57. Berlin: Mouton de Gruyter.
- Lieb, Hans-Heinrich & Sebastian Drude. 2000. Advanced glossing: A language documentation format. <http://www.mpi.nl/DOBES/documents/Advanced-Glossing1.pdf>.
- Matthews, P. 1974. *Morphology*. Cambridge: Cambridge University Press.
- Schultze-Berndt, Eva. 2006. Linguistic annotation. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 213-52. Berlin: Mouton de Gruyter.