

---

**The context of the context (and other factors to consider in describing a language)**

Christina M. Willis

---

Proceedings of Conference on  
**Language Documentation & Linguistic Theory 3**

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

19-20 November 2011 School of Oriental and African Studies, University of London

Hans Rausing Endangered Languages Project  
Department of Linguistics  
School of Oriental and African Studies  
Thornhaugh Street, Russell Square  
London WC1H 0XG  
United Kingdom

Department of Linguistics:  
Tel: +44-20-7898-4640  
Fax: +44-20-7898-4679  
linguistics@soas.ac.uk  
<http://www.soas.ac.uk/academics/departments/linguistics>

Hans Rausing Endangered Languages Project:  
Tel: +44-20-7898-4640  
Fax: +44-20-7898-4349  
elap@soas.ac.uk  
<http://www.hrelp.org>

© 2011 Christina M. Willis

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: 978-0-7286-0398-1

This publication can be cited as:

Christina M. Willis. 2011. The context of the context (and other factors to consider in describing a language). In Peter K. Austin, Oliver Bond, Lutz Marten & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory 3*, 307-315. London: SOAS.

or:

Christina M. Willis. 2011. The context of the context (and other factors to consider in describing a language). In Peter K. Austin, Oliver Bond, Lutz Marten & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory 3*. London: SOAS. [www.hrelp.org/eprints/ldlt3\\_32.pdf](http://www.hrelp.org/eprints/ldlt3_32.pdf)

## **The context of the context (and other factors to consider in describing a language)**

CHRISTINA M. WILLIS

*Department of Linguistics, Rice University*

### 1. INTRODUCTION

Increasingly, linguists interested in language description and documentation are turning away from exclusively employing a method of direct elicitation, which relies on translation from the contact language into the target language. Instead, the focus is on a ‘discourse-centered approach’ for data collection (Sherzer 1987, Urban 1991, and others), which aims at gathering natural speech. One argument for this method is that by focusing on naturally-occurring speech, the linguist can find and analyze words and structures that might not surface when sentences from one language are translated into another. Similarly, data obtained from speakers using their language in run-of-the-mill contexts will not suffer from unwanted effects of translation such as the literal interpretation of an utterance using the same structure found in the contact language. In this paper, I will explore some of the unusual and not always obvious consequences of using a single methodology in description and documentation work and advocate for a multi-faceted approach, which utilizes three methods of data collection while emphasizing naturally-occurring data.

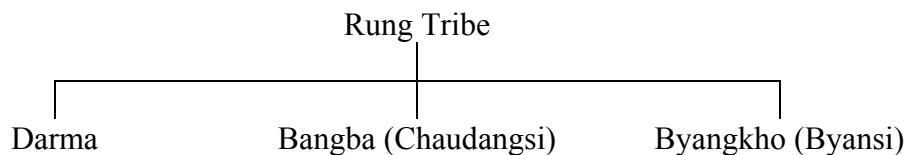
The examples for this discussion are drawn from my own fieldwork on an under-described Tibeto-Burman language spoken in the Indian Himalayas. While I followed the dictum, ‘record, transcribe, translate’, I also participated in local activities and spent much time interacting with and observing the speakers of the language I had set out to document and describe. It did not take long for me to discover that there were some texts that I could not adequately interpret without going beyond the immediate context in which they were recorded. In this vein, I will outline three issues that I encountered during fieldwork and data analysis. First, because not all structures can be elicited one must look at naturally occurring speech. Second, patterns of language use might not adhere to one’s expectations especially in communities that are multilingual. Third, sometimes the patterns found in natural-discourse need to be clarified in the context of an elicitation session.

This paper will be organized as follows. In Section 2, I will frame the context of my research by briefly outlining the linguistic scenario of my fieldwork site. Then in Section 3, I will reiterate the need for natural discourse, and present data from my research that reveal a pattern that would have gone unnoticed had I not recorded and analyzed natural speech. In Section 4, I will discuss the importance of understanding the local community when interpreting data. In Section 5, I will show how direct elicitation can help the field researcher clarify questions that arise during analysis of the texts. Finally, in Section 6, I will summarize the present discussion, highlight the implications of my argument with a focus on potential issues, and offer support for a recent trend in field research that includes a team approach to language description and documentation.

## 2. BACKGROUND

In 2001, as part of my doctoral training, I began a research project that focuses on a previously under-documented Tibeto-Burman language. The language, Darma, is spoken by approximately 1700 people in the Dharchula district of Uttarakhand, India. The speakers belong to a tribe, also called Darma, which is one of three local tribes identifying themselves as members of the macro Rung tribe (see Figure 1 below). Each tribe has a language associated with it and members of individual tribes report varying degrees of mutual intelligibility, which is correlated to the socio-economic factors of the local community (cf Willis to appear and Willis 2007). Many Rung people are in agreement that Darma is distinct enough from the other two Rung varieties to warrant its own description. As such, my research project was validated not just by the Darma people, but also by the broader Rung community.

**Figure 1**  
The tribes associated with the Rung macro tribe



The Dharchula area is bordered by Nepal and Tibet and is viewed by the Government of India as a sensitive area. It was restricted from the onset of the Indo-China conflict in the early 1960s until the early 1990s. An ongoing result of this conflict is the continued presence of several military groups in the area (e.g. border patrol, regional army, central intelligence, and so forth). Additionally, a handful of other inhabitants from outside the region live in Dharchula (e.g. missionaries and employees of agencies associated with a dam in the area). The outside population usually speaks Hindi or English when interacting with others in the community. The local population can be characterized as Rung or non-Rung. Members of the Rung tribe speak one or more Rung languages and one or more of the local non-Rung languages. The languages associated with the non-Rung population include a variety of Kumauni (which is reported to differ substantially from the under-described Eastern Kumauni) and a variety of Nepali. Most local inhabitants also speak Hindi and many local inhabitants also speak English; local schools are either Hindi- or English-medium. This wealth of languages used locally means that speakers have many options for language use in any given situation; I have observed that Rung people use the full array of linguistic resources.

## 3. RECORD, TRANSCRIBE, TRANSLATE

The call to focus on natural discourse as a source of linguistic data stems from the tradition in anthropology and linguistic anthropology that views language as the key element to understanding culture (Sherzer 1987, Urban 1991). That this method has

been adopted by linguists interested in describing languages is not surprising when one considers the wealth of data that can be amassed through recording, transcribing, and translating texts. One advantage to such data is its potential usefulness to other disciplines outside of linguistics (e.g. anthropology, folklore, history, and so forth). The value of such data is also clear in terms of attaining the goals of documentary linguistics which is ‘concerned with the making and keeping of records of the world’s languages and their patterns of use’ (Woodbury 2003: 35). As such, descriptive linguists are encouraged to base their analyses on texts of natural discourse and include sample texts in their grammars. In some instances scholars are encouraged to include texts with journal articles (e.g. the journal *Linguistics of the Tibeto-Burman Area*).

When preparing for fieldwork, I met with one of my committee members, Joel Sherzer, to talk about methods and strategies. My notes from that meeting centered on his directive to record, transcribe, and translate as much as I could while I was in the field. I followed his dictum to the best of my ability (and as circumstances would allow) and quickly discovered the advantages of doing so, which I will illustrate here.

During the early stages of my fieldwork I was not invited to record natural discourse by Darma speakers. They shied away from my recorder in part because I had yet to gain sufficient trust from key members of the Darma community. As such I did not have access to texts for analysis; thus I focused on direct elicitation with willing speakers. The existing sketches of Darma (Sharma 1989, Krishan 2001) described a split alignment system similar to Hindi and Nepali whereby the agents of past tense transitive constructions are overtly marked with ergative morphology. Indeed when I tested this with speakers, the analysis appeared to be accurate.

Eventually I was allowed to record people in their day-to-day interactions, and when I began analyzing these texts with the assistance of a native-speaker, it became apparent that the agentive marker was more widely distributed than had been reported. Example (1) below<sup>1</sup> illustrates that it is possible to mark the agent of a transitive verb in a non-past construction.

- (1) *hã niŋ sare suŋ su yeʃa da-da.*  
 then 1PL whole.LN village **ERG** guest do-3PL.NPT  
 ‘Then our whole village does the guests.’  
 Narrative describing how to propose marriage

Over time, I found agents overtly marked in past and non-past constructions with all persons and number. In one instance, the subject of an intransitive construction was marked as the agent of the action. This eventually led me to explore the role of the ergative marker, which may in fact have pragmatic functions beyond marking the

---

<sup>1</sup> The abbreviations used in this paper are: 1 = first person; 3 = third person; ABL = ablative; AUX.EX = existential auxiliary; COMPL = completive aspect; CVB = converb; DEM.NONVIS = demonstrative non-visible; ERG = ergative; HM = hesitation marker; LOC = locative; LN = loan word; NOM = nominalizer; NPT = non-past; PL = plural; PST = past; REL = relative.

agent of a transitive clause. Such patterns with the agentive marker are not unusual in other TB languages (see for example, Mongsen Ao as described in Coupe 2007, Kinnauri as described in Saxena 2007 and Kurtöp as described in Hyslop 2010).

The point I want to emphasize here is that without the natural discourse itself, I may not have discovered the pattern of the agentive marker as it is used in the day to day lives of Darma speakers. After realizing that it could appear in a wide variety of contexts, I sought to elicit the agentive marker from speakers in non-past constructions; it always required prompting (especially when the contact language was Hindi).

When we consider Darma in the context of South Asia, characterizing it as having a split-ergative system could have implications in terms of characterizing the area as a linguistic area. Furthermore, not understanding the patterns of agent marking that are possible in languages in general has implications for our understanding of its functions and uses and theories of ergativity.

#### 4. UNDERSTANDING THE COMMUNITY

When analyzing natural discourse it is important to recognize that the texts themselves are contextualized. A central criticism of descriptive linguists and their approach to language data pertains to the concept of a speech community (Himmelman 1998) and how it is routinely ignored. This is closely associated with the perception that those interested strictly in language description are likely to practice ‘helicopter linguistics’<sup>2</sup> meaning that they buzz in, gather data, and buzz out again. Typically, this is all done in a short period of time without any understanding of the local customs and practices. The criticism stems from the belief that researchers working on any language need to recognize that no community is homogeneous and the nuances of social organization must be attended to in order to understand the context of any given text (Hill 2003).

Similarly, the way discourse is organized in a given speech community varies and must be recognized for the researcher to be able to interpret texts (Keating & Egbert 2004). Additionally, researchers must recognize that they themselves are part of a community of practice in the role of researcher or language learner and that these very roles will have consequences for any situation where they are present (Hill 2003: 119-121). In my own research, I encountered a situation that left me puzzled for some time until I re-examined it from a local perspective.

Many Darma people wanted me to record a particular man who was reported to be an excellent story teller. I was able to record him early on in my research and indeed I found him riveting. I also understood quite a bit of what he was saying, which made me realize that much of the story he told was in Hindi. Upon analysis I discovered that in addition to using a lot of Hindi, he used the local variety of Kumauni and

---

<sup>2</sup> I attribute this term to Nora England. She used it in a panel discussion that we participated in at the Texas Linguistic Society Conference 10 held at the University of Texas at Austin in 2006.

Darma throughout the narrative. Despite this level of code-switching, some of the structures that he used are very difficult to elicit (e.g. the narrative use of the converb construction, which is used to move a storyline forward). Example (2) below illustrates the intricacies of the resulting text. In this example, you will notice loan words (cf ‘mischievous’, ‘mind’, ‘king’, ‘police’) and loan expressions (cf ‘like this’) as well as hesitation markers that are a blend of Hindi and Darma (cf *jo nini*) alongside the use of the converb construction.<sup>3</sup>

- (2) *to*      *?esa he*      *jo*      *?u*      *su*      *jo nini*  
 then      like.this.LN      that.LN      3SG      ERG      HM
- k<sup>h</sup>urapat*                      *Dimag*      *ga-lan*      *ju,*      *?u*      *su*  
 mischievous.idea. LN      mind. LN      do-CVB      after      3SG      ERG
- geʃ<sup>h</sup>o*      *me*      *po-lan*      *ju,*      *pulis*      *su*      *jo nini*  
 mill.flour      fire      light-CVB      after      police. LN      ERG      HM
- t'um-len*      *ju*      *jo nini,*      *raja*      *daro*      *sar*  
 capture-CVB      after      HM      king. LN      near      deliver
- pu-kur-su.*  
 COMPL-take.away-PST

‘Then it is like this, that he, that is, after planning the mischievous idea, after he set the mill on fire, the police, that is, after capturing (him), that is, (they) took him away and delivered (him) to the king.’

Story of Kiti Phondar

The text along with the overall context (i.e. I was recording a ‘good speaker’ of Darma telling a well-known story) illustrate that not every community, or individual, is going to have the same notion of what constitutes a ‘good speaker’. From my perspective, as a member of the linguistic community, being told that I was to hear a ‘good speaker’ left me with the impression that the orator would speak Darma without much code-switching. That this was not the viewpoint of the local community was something I did not immediately perceive. This serves to highlight my point, which is that the local ideology was neither what I expected nor what I perceived at the time I realized the extent of the code-switching in the text (i.e. that I had been duped). My ultimate interpretation, however, would not have been possible if I had not come to have a deeper understanding of the local community (and subsets therein). Nor would it have been possible if I had not returned to the text again and again wondering what else I could discover and puzzle over. This latter point of not leaving one’s data alone after a first analysis is addressed through the examples used

<sup>3</sup> The [tʰ] is a voiceless alveolar stop and the [t] is the voiceless dental stop.

in Hill (2003) as support for her integrated approach to language documentation. Sometimes a fuller understanding comes after puzzling over a text on multiple occasions over the course of time.

## 5. ELICITATION

I do not have the sense that any field linguist has deigned to eliminate elicitation as a method of data collection. In fact, elicitation is a side effect of any transcription and translation session that is a part of a discourse-centered approach to data collection. The concern that we would abandon elicitation as a thing of the dark ages is, however, salient. Here I will support the argument made by Chelliah & de Reuse (2011:12) that ‘the dialogue between elicitation and texts is crucial to the writing of a good descriptive grammar’. To illustrate this point, I will use an example of an elicitation session in which I discovered that a correlative construction found in natural discourse texts (as shown in example (3) below) was perceived to be a borrowed construction.

- (3) *ʔido*            *t<sup>h</sup>’ilo*            *t<sup>h</sup>’a-lan*    *bakte*    *ʔagar*    *ʃo*            *mi*  
 DEM.NONVIS    game.name    play-CVB    time    if.LN    REL.LN    person
- la-ru*            *su*    *ʔidu*            *t<sup>h</sup>’ilo*            *yu*    *la-n(i)*            *to*  
 hand.LOC    ABL    DEM.NONVIS    game.name    down    fall-3SG.NPT    then.LNS
- ʔidu*            *mi*            *dād*            *paʃ-ni*  
 DEM.NONVIS    person    fine. LN    must.LN-3SG.NPT
- ‘At the time of playing that game, if from some man’s hand that th’ilo falls, then that man must be fined.’
- Narrative about Cuti Gabla ceremony

I found that using Hindi as a contact language reproduced the same correlative constructions found in Hindi (e.g. ‘which girl is standing, that girl is tall’). These examples were all very similar to the example provided in (3) above. Using English as the contact language resulted in two sentences (e.g. ‘She took the baby into the house. Then, (she) kept it there.’). The Hindi examples were readily available in the natural discourse texts, but upon closer scrutiny, I noticed another structure. I found that I had to point to examples from texts of the third pattern and ask if my consultant could provide further examples, which he could do without hesitation. An example from one such instance is provided in (4) below. In this example, we find a nominalized verb that precedes the noun it modifies (literally: ‘the singing girl’).

- (4) *bera ga-nu t'eme filam su le.*  
 song do-NOM girl village.name ABL AUX.EX  
 'The girl who is singing is from Filam.'  
 Direct elicitation

During this session, my consultant indicated that the structure found in (4) was in fact 'better Darma' than the structure found in examples like (3).

The texts that contain the correlative construction have a few things in common. First, the relative-correlative markers are clearly loanwords. Second, the utterances in which these structures are found are often rich with other loan forms. In fact, one conversation that includes a fair amount of correlative constructions is a conversation about computers and technology, all of which is discussed using loans from English and Hindi. The variable patterns and the perception of each structure in terms of being 'indigenous' versus 'borrowed' would not have been discovered without the use of direct elicitation as a method.

## 6. CONCLUSION

It is my hope that the examples presented in this paper demonstrate that while natural discourse is imperative for adequate description and documentation, it should not be the only method that is used during fieldwork. We must also attend to information that can be gleaned through interview sessions and the direct observation and interpretation of speaking patterns. I argue that the goal for linguistic fieldwork should be the implementation of a three-pronged approach to data collection that uses naturally occurring speech as the basis for further exploration.

I would like to emphasize the importance of natural discourse, in part because I continue to see research projects and analyses that employ elicitation as a primary methodology. This is especially true for researchers interested in investigating a particular structure, but not interested in writing a grammar. But I would also like to draw attention to the importance of an integrated approach to methodology that is flexible and changes with each community where language description and documentation takes place.

Himmelman (1998: 164) charged that descriptive linguists have not paid enough attention to the methodology employed in data collection and analysis. This issue of methodology has been scrutinized since Himmelman argued that we must separate the descriptive endeavor from the documentary endeavor. One issue that relates to the separation of language description from documentation pertains to the goals of linguists and linguistic anthropologists, which are not incongruous and yet not identical. The former seeks to produce a grammatical analysis of the language while the latter aims to produce an ethnographic description. Both parties can accuse the other of falling short. The linguist does not include enough information about the cultural and social contexts, while the anthropologist does not include enough information about the language itself (e.g. analyzed texts). The integrated methodology approach that I advocate, is one which seeks to alleviate some of this



tension as the emphasis is on naturally occurring, culturally relevant discourse, which must be analyzed to be useful.

There are two issues that need to be considered when advocating for the use of multiple methodologies. The first issue that we must question is the lone field researcher approach. Woodbury (2003) has discussed the team approach to field research used in the Iquito Project, and I think that we should strive to make teamwork the norm. As we continue to discuss methodology and outline the expectations of the researcher, it becomes clear that the work will be more than many can handle. Moreover, the training will need to include exposure to both linguistic theory and anthropological theory. If we aim to recruit graduate students to join us in this important work of documenting and describing the languages of the world, we do not want to make the task less appealing by writing a job description that few can reasonably tackle. The second issue pertains to the presentation of data in the grammar. Using multiple methods of data collection means that the data needs to be presented in the final analysis using two techniques: (a) data must be identified in terms of its source and (b) the context of examples must be provided. Researchers will need to pay close attention to context and include a description of the recording session in their notes. Finally, a discussion of methodology needs to be included in grammars, papers, and presentations.

#### REFERENCES

- Chelliah, Shobhana Lakshmi & William De Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. New York: Springer.
- Coupe, Alexander R. 2007. *A Grammar of Mongsen Ao*. Berlin: Mouton de Gruyter.
- Hill, Jane H. 2005. The Ethnography of Language and Language Documentation. In J. Gippert, N.P. Himmelmann & U. Mosel (eds.), *Essentials of Language Documentation*, 113-28. Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus P. 1998. Documentary and Descriptive Linguistics. *Linguistics* 36, 161-95.
- Hyslop, Gwendolyn. 2010. Kurtöp Case: The Pragmatic Ergative and Beyond. *Linguistics of the Tibeto-Burman Area* 33, 1-40.
- Keating, Elizabeth & Maria Egbert. 2004. Conversation as a Cultural Activity. In Alessandro Duranti (ed.), *A Companion to Linguistic Anthropology*, 169-96. Malden, MA: Blackwell Publishing.
- Krishan, Shree. 2001. A Sketch of Darma Grammar. In Randy J. LaPolla (ed.), *New Research on Zhangzhung and Related Himalayan Languages*, 347-400. Osaka: National Museum of Ethnology.
- Saxena, Anju. 2007. Context Shift and Linguistic Coding in Kinnauri Narratives. In R. Bielmeier & F. Haller (eds.), *Linguistics of the Himalayas and Beyond*, 247-63. Berlin: Mouton de Gruyter.
- Sharma, Devi Datta. 1989. *Tibeto-Himalayan Languages of Uttarkhand*. New Delhi: Mittal Publications.

- Sherzer, Joel. 1987. A Discourse-Centered Approach to Language and Culture. *American Anthropologist* 89, 295-309.
- Urban, Greg P. 1991. *A Discourse-Centered Approach to Culture: Native South American Myths and Rituals*. Austin: University of Texas Press.
- Willis, Christina M. To appear. Documenting one language in a multi-lingual community. *Berkeley Linguistics Society* (BLS) 33.
- Willis, Christina M. 2007. *A Descriptive Grammar of Darma: An Endangered Tibeto-Burman Language*. Ph.D. Dissertation, UT at Austin.
- Woodbury, Tony. 2003. Defining documentary linguistics. In Peter K. Austin (ed.), *Language Documentation and Description, Volume 1*, 35-72. London: School of Oriental and African Studies.