

DIFFERENT TYPES OF DATA: A CASE STUDY OF GOEMAI DEMONSTRATIVES

Birgit Hellwig

Department of Linguistics, University of Erfurt, Germany
birgit.hellwig@uni-erfurt.de

This paper is set within the context of linguistic fieldwork on under-described languages across the world. It is concerned with methodological issues, and it explores different methods of data collection and the types of data they generate. The paper approaches this topic by means of a case study: the investigation of demonstratives in Goemai, a West Chadic language of Nigeria. This case study illustrates advantages and disadvantages of different methodologies, and the paper argues for combining methodologies when describing and documenting languages.

Cet article est mis dans le contexte de recherches sur le terrain linguistiques sur les langues à travers le monde. Il est concerné avec les questions méthodologiques, et il explore de différentes méthodes pour la collecte de données et les types de données qu'ils produisent. L'article s'approche de ce thème au moyen d'une étude de cas : l'investigation de pronoms démonstratifs dans Goemai, une langue tchadique du Nigeria. Cette étude de cas illustre des avantages et des désavantages de différentes méthodologies, et l'article argumente en faveur de la combinaison de méthodologies en décrivant et en documentant des langues.

0. INTRODUCTION

At the center of all language documentation is a focus on primary data. We are concerned with collecting and representing this data in such a way that it provides “a comprehensive record of the linguistic practices characteristic of a given speech community.” This includes both “the observable *linguistic behavior*, manifest in everyday interaction between the members of the speech community” as well as “the native speakers’ *metalinguistic knowledge*, manifest in their ability to provide interpretations and systematizations for linguistic units and events” (Himmelmann 1998:166).

This central concern with primary data inevitably leads to a focus on methods of data collection: there is an explicit interest in developing, refining and evaluating such methodologies. As a result, we have an increasing awareness of the advantages and disadvantages of different methods and, in particular, of the different kinds of data they generate. Relevant discussions are found in the more general literature on fieldwork methods within descriptive linguistics (e.g., Bouquiaux and Thomas 1987; Bowerman 2008; Crowley 2007; Everett and Sakel, to appear; Newman and Ratliff 2001; Samarin 1967; Vaux and Cooper 1999) as well as in the literature dedicated to language documentation (e.g., Gippert et al. 2006; Finnegan 2008; Gnerre 2008; Himmelmann 1998; Lüpke 2006; 2009; Seifart 2008). This paper is concerned with such fieldwork methodologies and their resulting data types, illustrating their interplay by means of the study of demonstratives in the West Chadic language Goemai.

There is general agreement in the above literature that elicitation allows researchers to assume control over the data: we can systematically check for known or presumed phenomena, fill gaps in our database, and ask speakers for their judgments about the acceptability and grammaticality of utterances. That is, elicitation gives us direct access to the speakers’ metalinguistic knowledge. But on the downside, elicitation sessions constitute an unusual and artificial genre. As a result, the elicited data may not represent the most natural way of saying something. Elicitation sessions are also found to be prone to misunderstandings between researchers and speakers, making it necessary to carefully check the reliability of the elicited data. Some of

these disadvantages can be counteracted by resorting to natural data: we gain access to the natural communicative practices of a speech community, can make informed judgments about the (un)naturalness and frequencies of different structures, and can discover new and unexpected phenomena. But these advantages of natural data have to be weighed against a number of disadvantages. It is labor- and time-intensive to create a comprehensive and representative corpus of natural data, and field-based corpora thus tend to be very small. As a result, they contain many accidental gaps and too few examples of most structures. Furthermore, natural data cannot contain negative evidence, i.e., we do not get information about whether or not an alternative expression could be used in a given context.

The literature therefore advocates a combination of different methodologies in order to capitalize on each method's advantages and to guard against its disadvantages. In addition, more and more researchers explore the uses of staged communicative events, i.e., of "communicative events that are enacted for the purpose of recording" (Himmelman 1998:185). This includes events that are staged by means of very general instructions ("please tell me a folktale"), but also more invasive procedures, e.g., the use of specific visual stimuli or props. Such staged communicative events strive to find a balance between control (as in the collection of elicited data), on the one hand, and naturalness (as in the collection of natural data), on the other.

This paper is concerned with such methodological issues: section 1 illustrates the contribution of different methodologies to the investigation of demonstratives in the Chadic language Goemai; and section 2 abstracts away from this case study and highlights some of its major points.

1. THE GOEMAI DEMONSTRATIVES

Goemai is a West Chadic language of Central Nigeria. There are approximately 200,000 Goemai, who live at the southern fringes of the Jos Plateau *sprachbund*, an old contact area between speakers of Chadic and Benue-Congo languages. Their long-lasting contact has led to considerable similarities in the lexical and grammatical structures of these non-related languages. Over the past decades, the pattern of language contact has changed, and the related Chadic language Hausa has emerged as the dominant lingua franca: all Goemai are today bilingual in Hausa, and this language is increasingly used as the main – and often only – language among the younger generations.¹

1.1. PREVIEW: AN ANALYSIS OF THE GOEMAI DEMONSTRATIVES

The Goemai demonstratives occur both as modifiers and as pronouns. As summarized in (1), they constitute morphologically complex words that contain a nominalizing prefix (optional in the case of the modifier, but obligatory in the case of the pronoun), an adverbializing prefix, a deictic classifier² and a deictic root. This

¹ I thank the many Goemai speakers who participated in this research, and especially the following people with whom I had long discussions about Goemai demonstratives: Louis Longpuan, Andreas Shakum, Thomas Longpuan, and Shalyen Mbai Nwang. I am very grateful to the Max Planck Institute for Psycholinguistic and to the Endangered Languages Documentation Programme for funding my research on Goemai.

² The term 'deictic classifier' is adopted from Aikhenvald (2000:176–183). Notice that classifiers are named after their morphosyntactic context, i.e., deictic classifiers occur in deictic expressions, but they classify nominals.

paper focuses on the deictic classifiers and the deictic roots (for the other two morphemes and for further details, see Hellwig, to appear:150–159).

(1) Goemai demonstratives³

<i>NMLZ</i>	<i>ADVZ</i>	<i>Deictic classifier</i>	<i>Deictic root</i>
gòe- (SG)	ń-	láng- (SG) ~ léng- (PL) ‘hang/move’	ń̀nòe ‘PROXIMAL’
mòe- (PL)		t'óng- (SG) ~ t'wót- (PL) ‘sit’	náng ‘DISTAL’
		d'yém- (SG) ~ d'yám- (PL) ‘stand’	
		t'ó- (SG) ~ t'óerép- (PL) ‘lie’	
		d'é- ‘exist’	

The deictic roots distinguish two grades of distance from the deictic center (see section 1.2), and the deictic classifiers specify the position of the referent (see section 1.3). The demonstrative word is used almost exclusively for exophoric purposes: it establishes joint attention to a referent within the physical surroundings, and it does not have any anaphoric functions (see section 1.4). The following sections 1.2 to 1.4 illustrate each aspect of the analysis in more detail, focusing on the underlying methodologies and data types. Section 1.5 then concludes this case study by suggesting avenues of future research.

1.2. DEICTIC ROOTS AND SPATIAL DISTANCE

The Goemai demonstratives distinguish two grades of distance from the deictic center: proximal and distal. The deictic center comprises the unit of speaker and addressee, and the proximal demonstrative is used for all referents that are close to this unit – independent of whether the referent is closer to the speaker or closer to the addressee. That is, it is used even when the referent is close to the addressee who himself is far away from the speaker (e.g., at the other end of a field). The proximal form is also used independently of whether or not the referent is within reach of one of the participants, or whether or not it touches one of them. The distal demonstrative, by contrast, is used for referents that are well beyond the reach of both the speaker and the addressee (e.g., at the other end of a field), including referents in large-scale geographical space (e.g., on the next hill). However, in the latter case, speakers preferably shift away from a demonstrative altogether, and instead use the deictic adverb **puánáng** ‘there/yonder’, i.e., the adverb takes over some of the demonstrative function.

The above paragraph has summarized the main factors that determine the distribution of the two deictic roots. The remainder of this section illustrates some of the underlying data types in more detail.

When starting fieldwork, the methodological possibilities are limited, and a fairly standard initial approach is to work with translations. And, indeed, most standard word lists ask for the translation equivalents of English ‘this’ and ‘that’ (e.g., the Swadesh list). In the case of Goemai, speakers invariably translate English ‘this’ with Goemai **ń̀nòe** or **ń-d'é-ń̀nòe** (arguing that “they mean the same”), and English ‘that’ with **ń-d'é-náng**. Alternatively, if Hausa is used as the metalanguage, they

³ I use an adapted version of the practical orthography developed by Sirlinger (1937). The following symbols may not be self-explanatory: **p'**, **t'**, **k'**, **f'**, **s'**, **sh'** = non-aspirated obstruents; **b'**, **d'** = implosives; **oe** = /ə/, **u** = /u/, **o** = /ɔ/.

In all example sentences, the free translation is followed by an identifier in brackets. This identifier links the example to the Goemai corpus archived with the Max Planck Institute for Psycholinguistics.

translate both Hausa **nàn** ~ **nân** ‘this (near speaker)’ and **nan** ‘that (near hearer)’ with Goemai **ń̀nòè** or **ń̀-đ'é-ń̀nòè**, and both Hausa **càn** ~ **cân** ‘that (distal)’ and **can** ‘that (remote)’ with Goemai **ń̀-đ'é-ń̀áng**. These responses immediately reveal an obvious problem with the use of translation equivalents: Goemai seems to have three expressions where English has two and Hausa four – and while it is likely that there is some overlap in their meanings, this overlap can only be partial. This problem is, of course, not restricted to Goemai. All books on fieldwork methodology warn against setting too much store by translation equivalents: the two expressions will never refer to exactly the same set of things in the real world, and they will have subtly different meanings. Ignoring Goemai **ń̀nòè** for the moment (see section 1.4), the demonstratives **ń̀-đ'é-ń̀nòè** and **ń̀-đ'é-ń̀áng** do not mean the same as English ‘this’ and ‘that’ (since, e.g., the English deictic center includes only the speaker – not the addressee). And they do not mean the same as their Hausa equivalents (which distinguish proximity to the speaker from proximity to the addressee, as well as two grades of remote distance).

At the beginning of fieldwork, it is, of course, not yet possible to state the meaning difference between an expression and its translation equivalent. It is, however, advisable to keep in mind that such equivalents can only ever approximate the language-internal meanings, and to resort to other data types. One possibility is to consult natural texts. In the case of the Goemai demonstratives, however, this alternative approach remains unsuccessful: the database of natural texts contains only very few demonstratives (see section 1.4 for the reasons), and – even more importantly – their reference context is often unknown. For example, the following sentence from a folktale does not give any information about, e.g., the spatial distance of the referent **sh'ùl** ‘shell, husk’ relative to the deictic center. Furthermore, these texts do not contain information about the equivalence of expressions, e.g., it is unknown whether or not the proximal demonstrative **ń̀-đ'é-ń̀nòè** of example (2) could be replaced with the distal form **ń̀-đ'é-ń̀áng** (and, if yes, with which meaning).

- (2) **Fuán yín, ògyà gwà góè f'ín sh'ùl**
 rabbit SAY nightjar SGM.LOG.AD.S OBLIG peel shell/husk

ń̀-đ'é-ń̀nòè.

ADVZ-CL:exist-DEM.PROX

The rabbit₁ said (to the nightjar₂), the nightjar₂ should peel this skin. (F99DLIGYA)

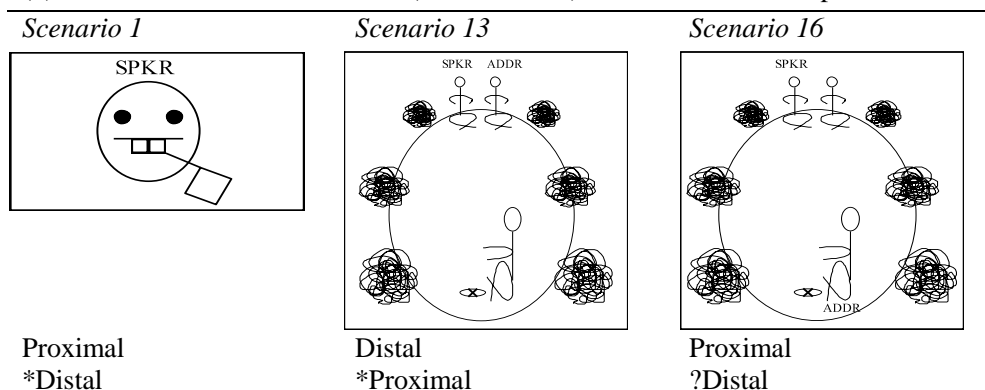
It is always possible to use natural text examples as the basis for further elicitation, e.g., to ask a speaker to illustrate the spatial distance of the referent in (2), or to discuss the alternative use of **ń̀-đ'é-ń̀áng**. This is a legitimate procedure, but it will not circumvent the fact that the database is constrained: given the few instances of demonstratives in natural texts, it is very likely that they will only illustrate an accidental subset of the relevant parameters. Furthermore, the reference context is unknown to the consultants, too: based on their native-speaker intuitions, they will be able to reconstruct possible contexts, but it will be difficult for the researcher to know which of the many circumstances of the reconstructed contexts are responsible for the choice of **ń̀-đ'é-ń̀nòè**.

A more promising approach is therefore to resort to some form of elicitation that systematically varies those parameters that are thought to govern the distribution of the two roots. For all forms of elicitation, it is necessary to have initial hypotheses that can then be verified or falsified by means of the elicited data. In the case of demonstratives, it is a reasonable assumption that they code spatial distance, i.e., that they anchor a referent in relative distance to a speaker and/or an addressee. This assumption is backed by the theoretical literature (e.g., Anderson and Keenan 1985;

Diessel 1999; Dixon 2003; Fillmore 1975; Halliday and Hasan 1976:31–87; Hanks 1992; Himmelmann 1996; Levinson 2004; C. Lyons 1999; J. Lyons 1977:646–677). Given the presumed central importance of the physical context, it is easily possible to structure the elicitation session around visual props. Against this background, a number of researchers have developed different staged communicative events. This section illustrates two such events that were of great importance for understanding the spatial semantics of the Goemai demonstratives.

The first investigation was based on a demonstrative questionnaire (Wilkins 1999), which proposes 25 different scenarios of referents anchored at different distances with respect to the speaker, the addressee and/or a third point of reference. These scenarios are then reenacted with real referents in real space, so as to match the depicted target scenario (see 3 below for some such scenarios). For example, the first scenario depicts a speaker pointing at his own teeth and saying “*this/that* tooth hurts.” We reenacted this scenario, and Goemai speakers would invariably choose the proximal form **ń-d'é-ńnòe**. I would then follow up on their spontaneous responses and ask whether alternative expressions can be used in reference to the same scenario. That is, I would check for negative evidence. In the case of scenario 1, all speakers rejected the distal form **ń-d'é-náng**. Conversely, speakers would choose the distal form but reject the proximal form in reference to scenario 13 – which pictures the speaker and addressee sitting next to each other, with the referent (a radio) at the other end of a football field, and the speaker saying to the addressee “*this/that* radio is a good one.” Scenario 16 is similar to scenario 13, but in this case, the referent is close to the addressee at the other end of the football field. Speakers always choose the proximal form for this scenario, but vary in their acceptance of the distal form: older speakers tend to reject it, while younger speakers tend to accept it.

(3) Some demonstrative scenarios (Wilkins 1999) and their Goemai responses



The second investigation was based on an interactive game termed “hidden color chips” (Enfield and Bohnemeyer 2001). For this game, the researcher places 16 objects in a large room or courtyard (following the manual’s instructions as to their placement), and hides differently colored chips under them. One speaker (the ‘memorizer’) observes the researcher, and his task is to remember which chips are hidden under which objects. A second speaker (the ‘checker’) arrives, and asks the memorizer about the locations of, e.g., the red chips. The memorizer then directs the checker to the correct referents, using the demonstratives and spatial expressions at his disposal. This game was designed as a follow-up to the questionnaire above: there, speakers often felt that more than one demonstrative was possible for a given scenario – but in actual use, speakers would, of course, only ever choose one. It was assumed

that an important factor that determines the choice of a demonstrative is attention-directing. The game was therefore designed to explore the use of demonstratives for directing the attention of addressees to referents in their physical surroundings. Example (4) is a typical instruction given by a Goemai memorizer. And example (5) happens to replicate aspects of scenarios 13 and 16 from the questionnaire (in 3 above). The memorizer directs the checker to open a cover. At first mention, both are several meters away from the referent, and the memorizer refers to it with the distal form. The checker then moves towards it and stops next to it, but does not show any other signs of having identified it as the intended referent. The memorizer therefore directs him again, and this time, he uses the proximal form – because the referent is now proximal to the addressee, and hence to the deictic center.

- (4) **Shàng góedé, uh, gòe-ń-d'é-ńnòe,**
 glance BOTTOM:GEN <HESIT> NMLZ(SG)-ADVZ-CL:exist-DEM.PROX
gòe=nà.
 2SGM.S=see
 Look under, uh, this one, you (will) see (the chip). (M01ANCOLOR)
- (5) **Èèp góe-ń-t'óng-náng puánáng=hòe. (...)**
 open(SG) NMLZ(SG)-ADVZ-CL:sit(SG)-DEM.DIST there/yonder=exactly
Á'à, gòe-ń-t'óng-ńnòe.
 no NMLZ(SG)-ADVZ-CL:sit(SG)-DEM.PROX
 Open that one over there. (...) No, this one. (M01ANCOLOR)

These two staged communicative events were instrumental in understanding the meanings of the deictic roots. On their basis, it was possible to develop the analysis presented at the beginning of this section. Crucially, both events provided contextual information: they made it possible for me to monitor and to manipulate real world scenarios (chosen so as to systematically test hypotheses that were taken from the theoretical literature on demonstratives). Access to this contextual information proved important for a number of reasons, as outlined below.

We know from the literature on fieldwork methodology that speakers always produce and judge expressions against a context. Therefore, if such a context is not provided, they will create it themselves – outside the control of the researchers. As a consequence, speakers and researchers may be unaware that they have different reference contexts in mind. And this fact is then responsible for misunderstandings. The two staged events above aim to prevent such misunderstandings by providing a clearly defined reference context for the demonstrative utterances. In the case of the demonstrative questionnaire, this aim was only partly successful. On the one hand, it allowed me to understand the distribution of **ń-d'é-ńnòe** and **ń-d'é-náng**. But on the other hand, the following problem emerged: speakers repeatedly referred to the same objects. This problem came up because I used the same referent in different scenarios, and because I asked follow-up questions about a referent within a single scenario. In all cases, the speakers were asked to imagine that they had just noticed the relevant referent for the first time, but this is not an easy task: it requires imagination, and it does not naturally reflect the current situation. As a result, this procedure can trigger misunderstandings and, indeed, speakers would often use the anaphoric form **ńnòe** instead of the exophoric forms **ń-d'é-ńnòe** and **ń-d'é-náng**. In retrospect, these responses do not pose a problem: I now know that **ńnòe** is an anaphoric demonstrative (see section 1.4), and I can reconstruct how my setup caused such a misunderstanding

to happen. At the time, however, these responses were problematic: I assumed that **ń̀nòè** constituted an exophoric demonstrative, and I tried in vain to understand how **ń̀nòè** differed from the other two forms – with speakers using them seemingly interchangeably, and explicitly saying that they mean the same.

It was possible to discover this misunderstanding through taking other data types into account. Wilkins (1999) already suggests to supplement the questionnaire data with data from participant observation, i.e., to pay attention to naturally-occurring situations that correspond to the questionnaire scenarios and to note down the spontaneous reactions of the speakers. In the case of Goemai, this procedure revealed that speakers would never spontaneously use **ń̀nòè** in such situations. These observations were confirmed by the results of the “hidden color chips” game. The advantage of this game over the observational data was its larger data set: for the observed events, the database was limited to those events that occurred when I happened to be present – i.e., the database is necessarily much smaller. And the advantage of the “hidden color chips” game over the questionnaire was its relative naturalness: the event was not overtly about demonstratives, but about finding hidden objects. Within this staged context, the speakers therefore used the demonstratives naturally. As a result, there was no problem in distinguishing between deictic and anaphoric uses. But despite this problematic aspect of the questionnaire, it made it possible to probe for alternatives (i.e., to discover what is and is not possible), and to engage in metalinguistic discussions with the speakers; plus it varied the relevant distance parameters much more systematically than the “hidden color chips” game did.

Access to a clearly defined reference context is also relevant for a second reason. Because the reference context is held constant, the two staged events can be run with a number of different speakers. This procedure makes it possible to compare responses and to thus introduce some measure of reliability. It also makes it possible to discover variation, both within languages and across languages. Within Goemai, it allowed me to discover that older speakers use the demonstratives differently from younger speakers (see the discussion of scenario 16 in 3 above; and also of examples 6 and 7 below). And across languages, a comparison of the results shows that the Goemai definition of the deictic center (as including both the speaker and the addressee) is typologically very unusual: a comparable deictic center is otherwise only known to exist in Brazilian Portuguese (Meira 2003; Meira and Guirardello-Damian 2002).

The staged communicative events above were designed to test specific hypotheses about the coding of distance within demonstratives, and about the use of spatial distance when directing attention to referents. For Goemai, this dedicated focus turned out to be both a strength and a weakness. Its strength was that it allowed me to discover how the Goemai demonstratives distribute with respect to referents anchored at various distances. Its weakness emerged when it was confronted with an unexpected situation: distance only plays a very marginal role within the Goemai demonstrative system. In most situations, Goemai speakers use the proximal form of the demonstrative, **ń̀-đ'é-ń̀nòè**. This emerged clearly from all data types considered so far: the questionnaire, the participant observation and the interactive game. Most questionnaire scenarios therefore elicited the same form, and the speakers felt this event to be repetitive and somewhat boring. One speaker gave a very telling response: after a number of scenarios that all elicited **ń̀-đ'é-ń̀nòè**, he refused to respond to the next scenario, and instead encouraged me to provide the answer myself – after all, he reasoned, it is very easy. And, of course, he was right: the appropriate response was,

yet again, **ń-d'é-ńnòe**. This anecdote highlights a central difference in expectations that is common to many field situations: the researchers' goals are to collect data that lends itself to a linguistic analysis, and many speakers' goals are to teach their language. In the latter case, success is measured by the researchers' progress in speaking the language. And the fact that I continued to ask obvious questions (i.e., questions that always elicited the same response) demonstrated my lack of proficiency, or – even worse – my unwillingness to learn the language.

The marginal importance of distance showed even more clearly in the data from the interactive game. In the questionnaire context, speakers would occasionally accept the distal demonstrative in small-scale space (provided that the referent was out of reach), but in the game, they would almost always use the proximal form. This includes situations such as (6) below, where there are two potential referents that are located at different distances (referent 9 is further away than referent 8). The questionnaire data shows that it would be possible to use the distal demonstrative in reference to referent 9, but speaker A. does not make use of this possibility. As a result, there is some confusion about the identity of the referent: speaker A. talks about referent 9, but speaker N. cannot decide between referents 8 and 9. In the end, the problem is only solved by speaker A. moving over and touching the intended referent. The speakers thus did not exploit the available two grades of distance. The only exceptions are younger speakers in their teens and twenties who are more likely to use the distance opposition to distinguish between referents in small-scale space (as in 7).

- (6) A: **Tákàrdá gòe-pyá=hók d'è ń-gòedè**
 paper NMLZ(SG)-become.white=DEF exist LOC-bottom:GEN
là=mótò n-t'óór góe ń-d'é-ńnòe.
 DIM(SG)=car(SG) LOC-flank 2SGM.POSS ADVZ-CL:exist-DEM.PROX
 The white paper is under the small car at this your side. [= referent 9]
- N: **B'ák góe-t'óór ń-d'é-ńnòe=hòe=à?**
 here PLACE-flank ADVZ-CL:exist-DEM.PROX=exactly=INTERR
Mótò góenàng dái? Góe-d'è góe-t'óór
 car(SG) which(SG) indeed NMLZ(SG)-exist PLACE-flank
ń-d'é-ńnòe=à?
 ADVZ-CL:exist-DEM.PROX=INTERR
 (Is it) here at this side? Which car now? (Is it) the one being at this side? [= referent 8 or 9] (M01ANCOLOR)
- (7) J: **Bì ń-d'é-ńáng t'óng k'à ńdè. (...)**
 thing ADVZ-CL:exist-DEM.DIST sit(SG) HEAD(SG):GEN one/other
Bì ń-d'é-ńnóe d'è k'á ńdè.
 thing ADVZ-CL:exist-DEM.PROX exist HEAD(SG):GEN one/other
 That thing sits on one. (...) This thing is on another one. [= referents 1 and 11] (Q01JCOLOR)

In the case of the questionnaire, the marginal importance of distance made it difficult to administer the task. And in the case of the game, it triggered confusion among the speakers as to the intended referents. This confusion prompted me to explore other avenues: the use of non-demonstrative strategies for uniquely identifying referents, and especially the coding of other semantic parameters within the demonstrative word. This latter point is taken up in the next section.

1.3. DEICTIC CLASSIFIERS AND POSTURAL SEMANTICS

Section 1.2 has shown that Goemai distinguishes two grades of distance from the deictic center – but also that this distance information does not play a major role for the use of the demonstratives: in most situations, speakers use the proximal form. While distance information thus tends to be neutralized, another type of information plays a more central role: postural information (coded in deictic classifiers). The demonstratives distinguish four postural categories (**láng-** ‘hang/move’, **t’óng-** ‘sit’, **d’yém-** ‘stand’ and **t’ó-** ‘lie’), plus one existential category (**d’é-** ‘exist’). The posturals give information about the orientation of the referent in space, and they are used to help the addressee identify the intended referent. The existential sets up a residual and default category whose distribution vis-à-vis the posturals is governed by pragmatic principles: it is only used if no postural is applicable, if the postural information cannot help identify the referent, or if the postural information is already coded elsewhere. That is, Goemai speakers and addressees identify referents largely on the basis of how they look (on the basis of their posture) – not of where they are (of their distance to the deictic center). This section presents the data types that made it possible to investigate the distribution of the deictic classifiers.⁴

The staged communicative events of section 1.2 vary the parameter of distance, i.e., they are concerned with a parameter that, cross-linguistically, is most likely to account for the distribution of demonstratives. Postural information, by contrast, is rarely coded in demonstratives. This phenomenon is only attested for a few languages: the Siouan and neighboring languages of North America (Barron and Serzisko 1982; Rankin 1977; 1988; Rood 1979), as well as the Guaykuruan languages of South America (Ceria and Sandalo 1995; Grondona 1998; Klein 1979; Vidal 1997). For Africa, there exist only some isolated reports: for the Nilo-Saharan language Mbay (Keegan 1997), for some Khoisan languages (cited in Kuteva 1999:204–205), and for the Chadic language Mada (Theda Schumann, p.c.). Given the infrequency of this phenomenon, it makes perfect sense that attention is not paid to posture when designing stimuli that are intended for comparative purposes. The stimuli of section 1.2 therefore did not generate much information on the distribution of deictic classifiers: posture was not systematically varied, and it did not help the addressee to identify the referent – speakers therefore almost always used the existential classifier (i.e., they tended to use the form **ń-d’é-ńnòè**). A similar situation obtains in translation and elicitation contexts: since postural information is irrelevant in such contexts, speakers would always volunteer forms with the existential classifier.

Nevertheless, the staged communicative events generated some isolated instances of postural classifiers. One such instance is found in example (5) above where the memorizer uses **t’óng-** ‘sit’ to distinguish the intended referent from a nearby ‘lying’ referent. Similarly, the natural texts contain a few examples of postural classifiers (as the form **láng-** ‘hang/move’ in 8).

- (8) **P’uàt dé gú=ná gòè-rwáng**
 exit(PL) SO.THAT 2PL.S:CONS=see NMLZ(SG)-become.mad
- ń-láng-náng t’óng sù (...)!**
 ADVZ-CL:hang/move(SG)-DEM.DIST IRR run(SG)
 Come out so that you see that moving mad person (who) would run (...)!
 (F00CGOERWANG)

⁴ It does not discuss the semantics of each classifier (for details of their semantic analysis and of the underlying methodology, see Hellwig 2006).

So far, the available information suggests that Goemai demonstratives code postural and existential information. However, there is not enough data available to investigate how such information is used in discourse. To investigate their distribution more systematically, I modified a staged communicative event developed by Levinson et al. (2001). They suggested that researchers collect and display a set of local objects, and that speakers are asked to discuss these objects (e.g., their relative merits, their uses etc.). The aim is to investigate the use of demonstratives in establishing and maintaining joint attention. In the case of Goemai, I followed their instructions, but additionally paid attention to posture: I varied the orientation for some objects, and held it constant for others. This was done with an array of local objects; and I also explored the uses of demonstratives in reference to surrounding landmarks and local flora.

On the basis of this kind of data, it was possible to observe that speakers provided postural information whenever they judged it relevant to the identification of the referent. For example, speakers exploited the postural information to distinguish between differently positioned calabashes in (9). In such cases, they would continue to use this postural information until the addressee has correctly identified the referent: in (10), speaker A. cannot immediately identify the referent after hearing the first demonstrative, and speaker N. therefore repeats the postural information and adds a descriptive phrase. By contrast, if this information cannot help identify the referent or if the referent can be identified on other grounds, the existential classifier will be used. For example, if the objects are in the same position, the postural information is irrelevant: speakers then preferably use the postural classifier in reference to the first object, but resort to the existential classifier in all other cases (as in the case of the three ‘sitting’ referents in 11).

- (9) **Gòe-ń-t'ó-ńnòe** **fà? (...)**
 NMLZ(SG)-ADVZ-CL:lie(SG)-DEM.PROX INTERR
T'ú **ń-t'óng-ńnòe=hòe=à?**
 calabash.bottle ADVZ-CL:sit(SG)-DEM.PROX=exactly=INTERR
 What about this lying one? (...) This sitting calabash bottle? (C01ANHAND)
- (10) N: **Gòe-ń-t'ó-ńnòe** **fà? (...)**
 NMLZ(SG)-ADVZ-CL:lie(SG)-DEM.PROX INTERR
Gòe-ń-t'ó-ńnòe **gòe-góe** **s'óng**
 NMLZ(SG)-ADVZ-CL:lie(SG)-DEM.PROX NMLZ(SG)-COMIT branch
ńnòe.
 LOC.ANAPH
 What about this lying one? (...) This lying one, the one with this handle.
- A: **Óhó,** **gòe-ńnòe** **à** **lúdè.**
 INTERJ NMLZ(SG)-LOC.ANAPH FOC calabash.spoon
 Oho, this is a calabash spoon. (C01ANHAND)

speech, i.e., they refer to entities present in the reported speech context. If the demonstratives had anaphoric functions, we would expect more examples to occur in the text corpus. And second, the corpus contains a handful of examples like (14) and (15), i.e., examples where a demonstrative co-occurs with another definiteness marker (the locative anaphor or the definite article). This co-occurrence suggests that the three words have different functions.

- (14) (...) **dóe** **póe,** **gùrùm** **ń-d'é-náng** **ńnòe.**
 ... come give person ADVZ-CL:exist-DEM.DIST LOC.ANAPH
 (...) (he) gave it to that (lit. that-existing this) man here. (#00CGOEBETLA)
- (15) **À** **wúròe** **t'ém** **tóe** **póe** **gòe** **dé** **gòe=éép**
 FOC who tell EMPH give 2SGM.O SO.THAT 2SGM.S:CONS=open(SG)
gwén **fridge** **ń-d'é-ńnòe=hòk** **yì?**
 ASSOC.PL fridge ADVZ-CL:exist-DEM.PROX=DEF CONS
 Who told you that you (can) open these (lit. these-existing the) fridges?
 (D00EWITCH3)

While the natural data points towards the above analysis, it does not constitute proof for it. There are simply too few demonstratives available to explore their relationship to the locative anaphor and the definite article. Furthermore, while they can co-occur with other definiteness morphemes, it is not clear, which morpheme contributes which meaning component. To explore these issues further, other data types are needed. In particular, the staged communicative events from section 1.3 proved useful for this purpose: they set up a discourse context that not only forces speakers and addressees to introduce and identify referents, but also to continue talking about the identified referents. Similarly, the “hidden color chips” game of section 1.2 triggers anaphoric uses once reference is established.

These data types show that the demonstratives are used for as long as the identity of the referent is questioned, asserted or corrected. This phenomenon is illustrated with the help of the dialogue in (16), which is taken from the “hidden color chips” game. Speaker N. uses a demonstrative and a pointing gesture to direct the addressee’s attention, but speaker A. is not sure whether he has identified the correct referent, and seeks confirmation by pointing to one of the potential referents and using a demonstrative. Once the referent is correctly identified and becomes background knowledge, the speakers then resort to the locative anaphor. This is illustrated in (17). Speaker A. first uses a demonstrative to seek confirmation that he has identified the correct referent, speaker N. confirms this, and speaker A. then uses the locative anaphor to talk about the color of the identified referent. The definite article =**hòk**, by contrast, cannot occur in place of the locative anaphor since it marks referents as identifiable from the discourse context, not from the spatial context.

- (16) N: **B'èp** **gòe=éép** **gòe-ń-t'óng-ńnòe.**
 do.again 2SGM.S=open(SG) NMLZ(SG)-ADVZ-CL:sit(SG)-DEM.PROX
 Open this sitting one again.
- A: **mm,** **gòe-ń-d'é-ńnòe=hòe=à?**
 yes NMLZ(SG)-ADVZ-CL:exist-DEM.PROX=exactly=INTERR
 Yes, this one?

N: **mm.**
 yes
 Yes. (M01ANCOLOR)

(17) A: **Gòe-ń-d'é-ńnòe=hòe=à?**
 NMLZ(SG)-ADVZ-CL:exist-DEM.PROX=exactly=INTERR
 (Is it) this one?

N: **mm.**
 yes
 Yes.

A: **Gòe-ńnòe, yìm lòn.**
 NMLZ(SG)-LOC.ANAPH leaf:GEN cloud
 This one is a blue paper. (M01ANCOLOR)

The various staged events again generated a large amount of data on the uses and non-uses of demonstratives. Without this data, it would not have been possible to prove the hypotheses emerging from the natural data: that Goemai demonstratives do not have anaphoric functions.

1.5. OUTLOOK

The above discussion of the Goemai demonstrative system has focused on three salient points: the semantics of the deictic roots (section 1.2), the contribution of the deictic classifiers (section 1.3), and the exophoric uses of the demonstrative word (section 1.4). There are additional aspects of this system, e.g., the contrastive use of demonstratives, or the gradual development of anaphoric functions of **ń-d'é-ńnòe** (i.e., of the form with the existential classifier and the proximal root). As in the other cases, these functions were investigated with the help of staged communicative events that were specifically designed to capture these parameters (for more information, see Hellwig, to appear:150–159).

This paper has concentrated on some central aspects of the Goemai demonstratives – and it has ignored other, equally central, aspects. For example, the demonstrative literature talks explicitly about demonstratives having gestural uses, i.e., it argues that demonstratives require a gesture and cannot be interpreted without monitoring physical aspects of the speech situation. This is true of Goemai, too. For example, reference to distal objects requires the use of the distal demonstrative – but apparently, it also requires the use of a gesture that takes the form of a fully stretched right arm, accompanied by a snapping of the fingers. This type of gesture is not attested in the case of proximal referents. But this statement is based only on impressionistic and anecdotal data. An adequate documentation of any demonstrative system would require an analysis of the accompanying gestures: Is pointing required? What form does the pointing take? How are the pointing gestures timed with respect to the spoken demonstratives? Do they mirror the distance and postural semantics coded in the demonstratives? Do they add information not coded in the demonstratives? The demonstrative questionnaire explicitly asks researchers to pay attention to these kinds of questions; and the staged communicative events are video-recorded, thus providing a data source for the gestural analysis. Another central aspect is intonation. It seems that the attention-directing properties of the Goemai demonstratives are reflected in their intonational properties: very often, they occur as pronouns, receive a rising intonation contour, and are set apart from the rest of the clause by a marked intonation break. Again, the audio- and video-recorded staged communicative events provide a

relevant data source that can be exploited for an intonational analysis. The literature on language documentation rightly draws our attention to the central importance of such factors in understanding spoken language (see, e.g., Seyfeddinipur, to appear, for the documentation of gesture, and Himmelmann 2006 for the documentation of intonation).

A further interesting aspect was already mentioned: all data types show that younger speakers use demonstratives differently from older speakers. That is, there seems to be a language change in progress: younger speakers are much more willing to use the distal demonstrative, and to exploit the proximal vs. distal opposition when drawing attention to referents. The attested data types are sufficient to state the existence of this difference. But they are not sufficient to investigate its reasons. Given that younger speakers are increasingly shifting away from Goemai towards Hausa, it is tempting to speculate that language contact and language attrition may be behind this change – but the sociolinguistic data that is necessary for such an investigation is missing.

2. CONCLUSION: TYPES OF DATA IN LANGUAGE DOCUMENTATION AND DESCRIPTION

This paper has presented a case study of the methodologies and data types used in the analysis of Goemai demonstratives. It is believed that at least some of the lessons of this case study have a wider application, and are valid to the investigation of other phenomena in a language, too. This section summarizes the relevant issues.

Of central importance were various types of staged communicative events that monitored the speech context and systematically varied the parameters of interest: distance, posture, establishing joint attention, and maintaining joint attention. These events are, in a way, similar to elicitation events: the researchers maintain a certain amount of control in that they manipulate the events in order to investigate specific hypotheses. But at the same time, such staged events have a major advantage over elicitation events: they provide all participants with systematic contextual information. This fact reduces the risk of misunderstanding, and it allows researchers to collect data from different speakers – thus making it possible to compare responses and to discover variation. At the same time, they need to be combined with elicitation: one crucial piece of information – negative evidence – can only be collected through eliciting speaker judgments as to whether or not an alternative expression (e.g., a different demonstrative) can be used in a given context.

Staged communicative events are not only similar to elicitation events. They also allow speakers to act naturally within the given context – a property that makes them more similar to natural events. They even have an advantage over such natural events in that the researchers are not dependent on the co-occurrence of the relevant events actually occurring spontaneously. In fact, Himmelmann (1998:186) explicitly argues that “[g]iven this somewhat pessimistic assessment of the possibilities of observing actual communicative events, it follows that one major concern of documentary linguistics in regard to data-gathering procedures will be with the evaluation and further elaboration of elicitation techniques and techniques for staging communicative events.” Such staged events allow the researchers to create relevant contexts themselves, thus being able to collect a large amount of data. But at the same time, these staged events cannot be used to the exclusion of spontaneously occurring events either: while speakers may talk naturally within a given context, the context itself is often artificial and may generate artificial data. If this is the case, their artificiality can only be discovered by comparing the data from staged events with the

data from natural events, looking for similarities and differences. Furthermore, both elicitation and staged communicative events test for known or presumed phenomena – the natural data, by contrast, draws our attention to unexpected phenomena.

All these considerations above were illustrated in the case study of section 1. Together, they served to investigate central aspects of the Goemai demonstrative system. They showed that the demonstratives code two grades of distance as well as postural information. They also showed how speakers use this information in actual discourse – discovering the primacy of postural information over distance information. And they showed how speakers use the demonstratives to draw the attention of addressees to referents in their physical surroundings. A reliance on either method (elicitation, staged events, natural data) to the exclusion of the others would not have produced these results.

ABBREVIATIONS

ADVZ	adverbializer	DIM	diminutive	LOG.AD	logophoric (addressee)
ANAPH	anaphor	EMPH	emphasis	M	masculine
ASSOC.PL	associative plural	FOC	focus	NMLZ	nominalizer
CL	classifier	GEN	genitive	O	object
COMIT	comitative	HESIT	hesitation	OBLIG	obligative
CONS	consequence	INTERJ	interjection	PL	plural
DEF	definite	INTERR	interrogative	POSS	possessive
DEM.DIST	distal demonstrative	IRR	irrealis	S	subject
DEM.PROX	proximal demonstrative	LOC	locative	SG	singular

REFERENCES

- Aikhenvald, A. Y. 2000. *Classifiers: A typology of noun categorization devices*, Oxford: Oxford University Press.
- Anderson, S. R., and E. L. Keenan. 1985. Deixis. In T. Shopen (ed.), *Language Typology and Syntactic Description 3*, 251-308. Cambridge: Cambridge University Press.
- Barron, R., and F. Serzisko. 1982. Noun classifiers in the Siouan languages. In H. Seiler and F. J. Stachowiak (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen 2*, 85-105. Tübingen: Gunter Narr.
- Bouquiaux, L., and J. M. C. Thomas. 1987. *Enquête et description des langues à tradition orale*, Paris: SELAF.
- Bowern, C. 2008. *Linguistic Fieldwork: A practical guide*, New York: Palgrave Macmillan.
- Ceria, V. G., and F. Sandalo. 1995. A preliminary reconstruction of Proto-Waikurian with special reference to pronominals and demonstratives. *Anthropological Linguistics* 37,2:169-191.
- Crowley, T. 2007. *Field linguistics: A beginner's guide* (edited and prepared for publication by N. Thieberger), Oxford: Oxford University Press.
- Diessel, H. 1999. *Demonstratives: Form, Function, and Grammaticalization*, Amsterdam and Philadelphia: John Benjamins.
- Dixon, R. M. W. 2003. Demonstratives. *A cross-linguistic typology*. *Studies in Language* 27,1:61-112.
- Enfield, N. J., and J. Bohnemeyer. 2001. Hidden colour-chips task: Demonstratives, attention, and interaction. In S. C. Levinson and N. J. Enfield (eds.), 'Manual' for the field season 2001, 21-28. Nijmegen: MPI.
- Everett, D. L., and J. Sakel. To appear. *Linguistic fieldwork: A student guide*, Cambridge: Cambridge University Press.
- Fillmore, C. J. 1975. *Santa Cruz Lectures on Deixis 1971*, Bloomington: Indiana University Linguistics Club.
- Finnegan, R. 2008. Data – but data from what? In P. K. Austin (ed.), *Language Documentation and Description 5*, 13-28. London: SOAS.

- Gippert, J., N. P. Himmelmann, and U. Mosel, eds. 2006. *Essentials of language documentation*, Berlin and New York: Mouton de Gruyter.
- Gnerre, M. 2008. Linguist's multi-layered data and the linguistic community's polyphony. In P. K. Austin (ed.), *Language Documentation and Description* 5, 29-59. London: SOAS.
- Grondona, V. 1998. Location and direction in Waikurúan languages. *Berkeley Linguistics Society* 24:106-116.
- Halliday, M. A. K., and R. Hasan. 1976. *Cohesion in English*, London: Longman.
- Hanks, W. F. 1992. The indexical ground of deictic reference. In A. Duranti and C. Goodwin (eds.), *Rethinking Context: Language as an interactive phenomenon*, 43-76. Cambridge: Cambridge University Press.
- Hellwig, B. 2006. Field semantics and grammar-writing: Stimuli-based techniques and the study of locative verbs. In F. K. Ameka, A. Dench and N. Evans (eds.), *Catching language: The standing challenge of grammar writing*, 321-358. Berlin and New York: Mouton de Gruyter.
- Hellwig, B. To appear. *A grammar of Goemai: A West Chadic language of Nigeria*, Berlin and New York: Mouton de Gruyter.
- Himmelmann, N. P. 1996. Demonstratives in narrative discourse: A taxonomy of universal uses. In B. A. Fox (ed.), *Studies in Anaphora*, 205-254. Amsterdam and Philadelphia: John Benjamins.
- Himmelmann, N. P. 1998. Documentary and descriptive linguistics. *Linguistics* 36,1:161-195.
- Himmelmann, N. P. 2006. Prosody in language documentation. In Gippert et al., 163-181.
- Keegan, J. M. 1997. *A reference grammar of Mbay*, München and Newcastle: Lincom Europa.
- Klein, H. E. M. 1979. Noun classifiers in Toba. In M. Mathiot (ed.), *Ethnolinguistics: Boas, Sapir and Whorf revisited*, 85-95. Den Haag and Paris: Mouton.
- Kuteva, T. A. 1999. On 'sit'/'stand'/'lie' auxiliations. *Linguistics* 37,2:191-213.
- Levinson, S. C. 2004. Deixis and pragmatics. In L. R. Horn and G. Ward (eds.), *The Handbook of Pragmatics*, 97-121. Oxford: Blackwell.
- Levinson, S. C., S. Kita and A. Özyürek. 2001. Demonstratives in context. In S. C. Levinson and N. J. Enfield (eds.), 'Manual' for the field season 2001, 52-54. Nijmegen: MPI.
- Lüpke, F. 2006. Small is beautiful: Contributions of field-based corpora to different linguistic disciplines, illustrated by Jalonke. In P. K. Austin (ed.), *Language Documentation and Description* 3, 75-105. London: SOAS.
- Lüpke, F. 2009. Data collection methods for field-based language documentation. In P. K. Austin (ed.), *Language Documentation and Description* 6. London: SOAS.
- Lyons, C. 1999. *Definiteness*, Cambridge: Cambridge University Press.
- Lyons, J. 1977. *Semantics*, Cambridge: Cambridge University Press.
- Meira, S. 2003. 'Addressee effects' in demonstrative systems: The cases of Tiriyo and Brazilian Portuguese. In F. Lenz (ed.), *Deictic conceptualisation of space, time and person*, 3-12. Amsterdam and Philadelphia: John Benjamins.
- Meira, S. and R. Guirardello-Damian. 2002. Non-contrastive exophoric use of demonstratives in spoken Brazilian Portuguese. Unpublished manuscript.
- Newman, P., and M. Ratliff (eds.). 2001. *Linguistic Fieldwork*, Cambridge: Cambridge University Press.
- Rankin, R. L. 1977. From verb to auxiliary to noun classifier and definite article: Grammaticalization of the Siouan verbs 'sit', 'stand', 'lie'. In R. L. Brown, K. Houlihan and A. MacLeish (eds.), *Mid-America Linguistics Conference*, 273-283. Minnesota: University of Minnesota.
- Rankin, R. L. 1988. Quapaw: Genetic and areal affiliations. In W. Shipley (ed.), *In honor of Mary Haas: From the Haas Festival Conference on Native American Linguistics*, 629-650. Berlin, New York and Amsterdam: Mouton de Gruyter.
- Rood, D. S. 1979. Siouan. In L. Campbell and M. Mithun (eds.), *The languages of native America: Historical and comparative assessment*, 236-298. Austin and London: University of Texas Press.
- Samarin, W. J. 1967. *Field linguistics: A guide to linguistic field work*, New York: Holt, Rinehart and Winston.
- Seifart, Frank. 2008. On the representativeness of language documentations. In P. K. Austin (ed.), *Language Documentation and Description* 5, 60-76. London: SOAS.
- Seyfeddinipur, Mandana. To appear. Some reasons for documenting gestures. In N. Thieberger (ed.), *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press.
- Sirlinger, E. 1937. *Dictionary of the Goemai language*. Unpublished manuscript.
- Vaux, B., and J. Cooper. 1999. *Introduction to linguistic field methods*, München and Newcastle: Lincom Europa.
- Vidal, A. 1997. Noun classification in Pilaga: Guaykuruan. *Journal of Amazonian Languages* 1:60-111.
- Wilkins, D. 1999. Demonstrative Questionnaire: "this" and "that" in comparative perspective. In D. Wilkins (ed.), 'Manual' for the 1999 field season, 1-24. Nijmegen: MPI.