

Language Documentation and Description

ISSN 1740-6234

This article appears in: *Language Documentation and Description*, vol 7. Editor: Peter K. Austin

Research methods in language documentation

FREDERIKE LÜPKE

Cite this article: Frederike Lüpke (2010). Research methods in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, vol 7. London: SOAS. pp. 55-104

Link to this article: <http://www.elpublishing.org/PID/082>

This electronic version first published: July 2014



This article is published under a Creative Commons License CC-BY-NC (Attribution-NonCommercial). The licence permits users to use, reproduce, disseminate or display the article provided that the author is attributed as the original creator and that the reuse is restricted to non-commercial purposes i.e. research or educational use. See <http://creativecommons.org/licenses/by-nc/4.0/>

EL Publishing

For more EL Publishing articles and services:

Website:	http://www.elpublishing.org
Terms of use:	http://www.elpublishing.org/terms
Submissions:	http://www.elpublishing.org/submissions

Research methods in language documentation

Friederike Lüpke

An increasing interest in linguistics may be noted among workers in anthropology, culture history, sociology, psychology, and philosophy. For all of them linguistics is of basic importance: its data and methods show better than those of any other discipline dealing with socialized behavior the possibility of a truly scientific study of society. Linguists should, on the other hand, become aware of what their science may mean for the interpretation of human conduct in general. (Sapir 1929: 207)

1. The new role of data

Language documentation, understood as the creation of corpora of annotated and translated speech data in audio and video format, is a recently developed field of empirical linguistics (Himmelmann 1998, 2006a; Woodbury 2003). Since it owes its existence partly to rapidly evolving digital technologies for the recording, processing, and archiving of these data, it is not surprising that the main methodological focus so far has been on the form a documentation should take. Thus, we find explicit recommendations on such things as suitable data formats, form and content of metadata descriptions, minimally required levels of annotation, and access to and portability of data (Austin 2006; Bird & Simons 2003; Nathan & Austin 2004; Nathan 2006; Thieberger 2004; Wittenburg et al. 2002). This paper focuses on another central question for good practices in language documentation and description (henceforth LDD): what linguistic goals it can have and how methods of data collection are related to achieving these goals. By doing so, it joins a growing body of literature drawing attention to the various methodological challenges of LDD. These issues include:

- translation and translatability and lexical knowledge (Bradley 2007; Evans & Sasse 2007; Haviland 2006; Woodbury 2007);
- adequate documentation of semantics (Hellwig 2006a, b; Matthewson 2004)
- documenting pragmatics (Bergqvist 2007; Grenoble 2007);
- guidelines for the documentation of phonetics (Ladefoged 2003);

- the data necessary for investigating prosody (Himmelman 2006b);
- grammar and dictionary writing (Ameka et al. 2006; Mosel 2004, 2006; Payne & Weber 2007);
- sample size and structure (Lüpke 2005a; Seifart 2008)
- language and linguistic ideologies influencing research practice (Barwick 2005; Foley 2003, 2005);
- inter-disciplinarity and multi-disciplinarity of the field (Barwick 2005; Eisenbeiss 2005; Finnegan 2008; Franchetto 2006; Harrison 2005; Hill 2006; Widlok 2005);
- the advantages and challenges of teamwork (Dwyer 2006; Kibrik 2006);
- usability of the corpus and description for the speech community (Florey 2004; Mosel 2004; Nathan 2006; Seifart 2006; Woodbury & England 2004).

Rather than concentrating on a particular area of language or linguistics, in this paper I adopt a bird's eye view on a documentary corpus and present a typology of data collection methods.

Since language documentation has data at its very core, it is important to extend this focus on data to the ways in which they are collected, if language documentation is to fulfil its goals (see also Nathan 2009). These are generally understood to be the creation of a record of the actual linguistic practices of a speech community¹ designed for a broad audience ranging from linguists and researchers from neighbouring disciplines to members of the speech communities whose languages are documented (Austin 2006; Austin & Grenoble 2007; Himmelman 1998, 2006a; Woodbury 2003). Already, the goal of observing and documenting language use (*parole*) distinguishes LDD from linguistic frameworks concentrating on I-language, competence, or *langue* as a non-observable phenomenon. This has consequences for the nature of the data to be collected as well as their interpretation.²

¹ My use of the term 'speech community' follows Hymes's definition: "a community sharing rules for the conduct and interpretation of speech" (Hymes 1972: 54).

² While seemingly trivial, this distinction between goals and the consequent nature of the data is often neglected in cross-theoretical and cross-linguistic comparisons (see also Haspelmath 2007 *pace* Newmeyer 2007). In the following, I limit myself to a discussion of data from the perspective of usage-based or functionalist approaches to linguistics.

LDD corpora crucially need to integrate analysis and hence cannot be limited to observed and observable linguistic behaviour but have to complement data that reflects this behaviour with information on negative evidence, metalinguistic awareness, felicity conditions for utterances, etc. Thus, the corpora need to contain data resulting from a battery of methods aimed at elucidating different aspects of linguistic structure and knowledge. A documentary corpus consequently needs to be as representative and varied as possible, not only in terms of what social variables are covered through the choice of speakers of different ages, genders, etc. but also in terms of the linguistic parameters and data types included. This is what this paper attempts: an overview of different methods of linguistic data collection in field-based documentary research to illustrate how different ways of gathering data have an influence on the insights researchers can obtain from them, and to explore how different kinds of data contribute jointly to providing insights into the structure of a language, the linguistic intuitions of its speakers, and their repertoires.

2. General design considerations for documentary corpora

Documentary corpora generally have two purposes:

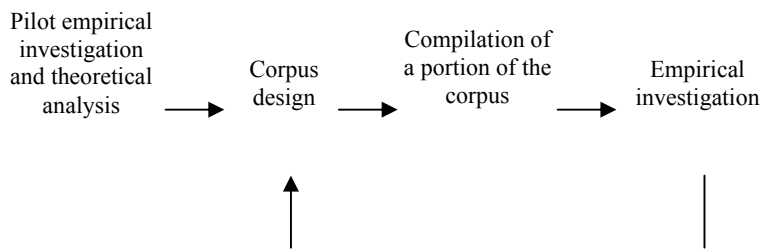
- to serve the specific research goals of a particular project by providing the data base for particular linguistic/ethnographic/historical analysis;
- to capture the language use of a community as comprehensively as possible.

These purposes, although intimately related, are potentially in conflict. Additionally, field-based corpora often constitute first documentations, where it is impossible to define clear research goals prior to the first stage of data collection. Therefore, a field-based corpus needs flexibility to allow for emerging hypotheses and changing interests in the course of a project, while including some degree of representativeness as a record of the practices of the targeted speech community. Even if field-based corpora necessarily lack the dimensions and degree of computational sophistication of large corpora of major languages, an awareness of design considerations from the earliest stages of project planning onwards is crucial.

2.1 Cyclic corpus design

Biber (1993: 256) describes corpus linguistics workflow as a cyclic activity with the following schematic stages:

Figure 1: Cycle of corpus work after Biber (1993)



Different considerations guide the parts of the cycle. In the following, I limit myself to a discussion of those particularly relevant for LDD.

Initially, the target population and its boundaries need to be defined. In LDD, this involves identification of the speech community, and its linguistic production in terms of extralinguistic (or situational) and linguistic characteristics. The former is labelled as genres/registers, and the latter as text types by Biber (1993). Such labelling presupposes an inventory of genres/registers and information on homogeneity and variation across text types and speakers that can in principle only be the result of corpus investigation. Initial hypotheses on the population therefore need to be carefully assessed in later stages of corpus work, and the pilot corpus design adjusted if necessary.³ Rather than choosing a demographic sample and representing speech production proportionally, it is advisable to represent the full range of situational features, i.e. genres and registers, regardless of their rarity of occurrence in the population, and to stratify the sample carefully rather than using probabilistic sampling techniques. This is reflected in the relative importance of low-frequency genres and registers (e.g. formal speeches or minutes of a departmental meeting) as opposed to conversation, the genre making up the bulk of everyday language use. Stratified sampling also ensures that linguistic features associated with rarer text types are represented in sufficient quantity (e.g. passives, which are rare in unplanned speech, but more frequent in planned speech – see Biber 1995; Ochs 1979). Of course, good intuitions on the linguistic repertoire of the speech

³ Biber (1995) suggests a number of statistical tests to assess the representativeness of a corpus by calculating the necessary sample size and measuring variation. However, he himself concedes that these tests are often circular, since they are based on the standard deviation in a sample that is used as the benchmark and hence on the representativeness of that sample. Therefore, it seems that intuition needs to be a guiding factor in corpus design and sample selection.

community are indispensable in order to arrive at a stratified sample, and it is recommended to conduct a detailed sociolinguistic investigation, comprising attitude statements, judgment tasks, and observation of communicative practices, together with collection of demographic information on the population, prior to corpus design.

2.2 Frequency considerations

High-frequency linguistic features are well-represented even in small samples – Biber demonstrates robust distribution of a number of features in 1,000 word samples – but rare features require much larger samples. Therefore, it is recommended that the sample be supplemented with carefully chosen texts containing low-frequency phenomena. In LDD, these phenomena can often only be documented through the inclusion of data resulting from staged communicative events and elicitation (see 5 and 6 below). The distribution of linguistic features within texts also deserves some consideration: while robust features are distributed evenly within texts and can be captured through short segments, rare features exhibit much more variability and can only be captured through the inclusion of longer texts. In addition, the ‘curvilinear’ nature of many features – the fact that some features occur more frequently at the beginning of texts – needs to be taken into account by aiming at a high diversity across texts.

2.3 Variation within the corpus

Sample size is particularly relevant when assessing the degree of variation within and across genres and registers, i.e. their common and differentiating linguistic features and the extent of their variation. Biber (1990) shows that for frequent features, a 10-text sample gives a good overview of the characteristics of a register. However, rare features are distributed much more unevenly, and more sophisticated techniques must be used to determine variation and sample size necessary to represent it (see Biber 1990, 1993, 1995; and also Bernard 2002, chapters 6-8 on sampling methods in anthropology). For LDD corpora of endangered languages, attempts to reach an optimal representation of variation may be futile in light of limited spheres of use and small speaker numbers.

2.4 Longitudinal versus cross-sectional design

Documentary corpora aim at being open so data can be added at any time (what Woodbury 2003 calls ‘opportunistic’). If the advantages of this openness are fully exploited, the corpora could be used to investigate

diachronic variation by integrating into the corpus design longitudinal aspects, i.e. samples of language use taken over a period of time. Longitudinal corpora are commonly used in language acquisition research, and helpful guidelines are available (cf. Eisenbeiss 2005). Alternatively, a corpus can be intended to be cross-sectional, offering a snapshot of language use at a given time. While few documentary corpora explicitly use a longitudinal design, the fact that the data may result from several stages of fieldwork and data collection, often incorporating legacy materials, means that the corpus may already include some time-depth. Explicit, documented design of the corpus along with well-structured metadata may facilitate future longitudinal studies.

2.5 Multilingual and multilectal documentation

Many field-based corpora focus on representing language use in one variety, however defined. However, LDD research often takes place in multilingual and multilectal settings in contexts where no standard variety exists and documentation of other languages/varieties spoken by the speech community is unavailable. The lack of standardised varieties adds another dimension of complexity to the documenters' task of defining a variety or varieties to be represented in the documentation, because speakers can be expected to associate different varieties to particular domains of language use, and to routinely borrow from other varieties they speak or to code-switch between them. In addition, it is often impossible to differentiate between borrowing and code-switching in the absence of a detailed sociolinguistic and linguistic investigation. While it is often easy to detect influence from a language not closely related to the one under investigation, closely related varieties (be they labelled as distinct languages or as dialects) are notoriously difficult to identify. Yet, how to deal with the inherently hybrid nature of language use in multilingual and multilectal communities is increasingly recognised as important in LDD, in view of the diglossic or triglossic situations of language use that prevail. These should be adequately documented since they are part of the linguistic practices of a community. In addition, there is a growing awareness of the importance of geographical factors and language contact in shaping language structure (Haspelmath et al. 2005; Heine & Nurse 2008; Matras et al. 2006). LDD corpora can contribute the data needed to investigate these topics. Therefore, rather than concentrating on a monolingual corpus (that relegates contact phenomena to, for instance, a section on loanwords in the description), a number of ongoing LDD projects aim at representing all varieties spoken by a given speech community or in a given geographical area. Whatever the scope of a field-based corpus, the question of how to handle multilingualism needs to be addressed, since the majority of the world's languages are spoken in bilingual or multilingual speech communities, and all languages exhibit dialectal and sociolectal variation.

Including multilingual and multilectal speech situations can result in a proper representation of all the genres and registers attested in the varieties under investigation.

2.6 Individual-language versus cross-linguistic focus

For a number of larger and well-described languages, there exist parallel corpora (i.e., corpora incorporating a text in one language and its close translation equivalent in another language). In multilingual and multilectal speech situations (mentioned in 2.4 above), it is important to consider whether part of the corpus should provide ‘parallel texts’ in order to create data that are immediately comparable. For observed communicative events (see 4 below) this is probably only possible for languages spoken in at least partly similar cultural settings. For staged communicative events, however, it is possible to collect directly comparable data from a wide range of languages; there is a large and expanding number of corpora including, e.g. renderings of the ‘Pear story’ (Chafe 1980) or the ‘Frog story’ (Berman & Slobin 1994). Likewise, a number of questionnaires such as Dahl (1985) on tense-aspect, have resulted in a huge body of ‘parallel’ elicited and staged communicative events (see 5 and 6 below). Inclusion of questionnaires in an LDD corpus allows direct exploitation of the data for typological purposes and links them to areas of cross-linguistic interest. Therefore it is useful to incorporate considerations of cross-linguistic comparability into the corpus design.

2.7 Negative evidence

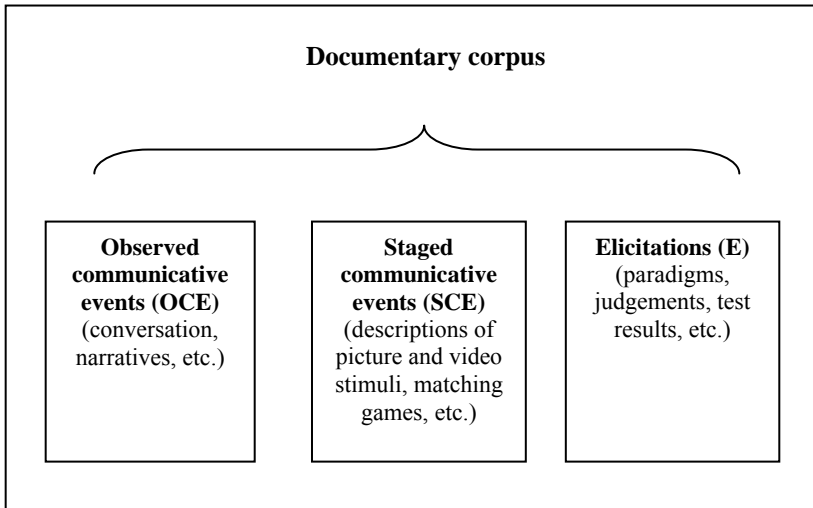
At present, there is a great emphasis on the collection of speech usage data in LDD, especially because for endangered languages these data may constitute the last or only record of a language. This phenomenological focus has one important limitation, however: LDD corpora (unlike corpora of languages with long descriptive traditions) cannot exist independently of linguistic analysis. Such analysis needs to rely on an additional type of evidence which is not directly observable, namely negative evidence, i.e. information on items and constructions or contexts and felicity conditions that are ruled out in the investigated variety. For example, it is impossible to conclude from the absence of *writed* in even the largest corpus of English that this form is ungrammatical; rather, it might simply be accidentally unattested in the corpus. Given the size limitations of most field-based corpora, some thought should be given to obtaining and representing the negative evidence necessary to arrive at descriptive generalisations. We address the design and methodological consequences of this in the following section.

3. Types of communicative events resulting from different methods of data collection

As emphasised above, LDD corpora ideally contain representative samples of language use of a given speech community complemented by data providing negative evidence and insight into linguistic intuitions. It is useful to have a way of categorizing these different types of data based on the manner in which they were collected. Himmelmann (1998) presents a three-way distinction of ‘communicative events’, as he calls them, which are ideally included in a field-based documentary corpus. He distinguishes the following event types (see Figure 2):

- Observed communicative events (OCEs) where the only influence of the researcher is (ideally) their presence;
- Elicitations (Es) being communicative events heavily influenced linguistically by and only created for the sake of the researcher, such as word lists, paradigms or acceptability judgments.
- Staged communicative events (SCEs) occupy a middle ground between OCEs and Es: they are prompted or ‘staged’ for linguistic purposes, but often use non-linguistic prompts such as pictures, video clips that consultants are asked to sort or describe, or games they are invited to play and describe. While they owe their existence to the research project and hence do not constitute speech events in the sense of Hymes (1972), their linguistic structure is less likely to be directly influenced by the researcher than that of elicitations.

Figure 2: Communicative events in a documentary corpus, after Himmelmann (1998, 2006a)



The different event types and their subtypes do not yield equivalent data. Ideally, an LDD corpus will balance the different types collected with a range of speakers to create a maximally informative and diverse corpus. In addition, researchers can use different types of events to complement each other for analytical purposes, as is custom in the social sciences, where triangulation is a commonly practiced research method. Triangulation requires consideration of the advantages and disadvantages of data resulting from these event types; in the following sections we explore their usefulness for linguistic analysis (OCEs in 4, SCEs in 5, and Es in 6).

OCEs have long occupied a prominent position among the data collected by field linguists, and a large body of specialised literature on them exists⁴. Likewise, detailed recommendations for elicitations are available elsewhere. Therefore, these two event types will only be discussed globally; the main focus in this paper will be to present the first systematic overview of SCEs.

⁴ Research schools concerned with OCEs of different types and their collection and analysis comprise ethnography of communication, discourse and conversation analysis, register and genre studies, and frameworks investigating oral history and verbal art, among others. While it is impossible to give a selection of relevant references for all these fields, some of them will be mentioned in the text.

4. Data resulting from observed communicative events

There is no universal way of categorizing ‘natural’ speech events or OCEs. Therefore, a good heuristic to capture the repertoire of a speech community is to aim at including those speech events that are recognised by its members (for instance by having a name in the language), and to characterise these events in terms of parameters employed in the ethnography of speaking, corpus linguistics and language documentation. Hymes’s (1972) acronym ‘SPEAKING’, classifies a speech event in terms of its **S**etting, **P**articipants, **E**nd or purpose, **A**ct sequence, **K**ey or tone, **I**nstrumentalities or channels, **N**orms of interaction and interpretation and **G**enre. Parameters listed by Biber (1993) and Himmelmann (1998), and partly reflected in metadata standards (e.g. in the IMDI set), draw on Hymes’s criteria and add a number of others. Himmelmann employs the parameter of plannedness, which has been shown to have important linguistic consequences (Ochs 1979), to create a cline of OCEs, as set out in Table 1.

Table 1: OCEs as a cline from unplanned to planned, after Himmelmann (1998: 180)

Parameter	Major types	Examples
unplanned	exclamative	‘ouch!’ ‘fire!’
	directive	‘scalpel!’ greetings
	conversational	small talk chat
		discussion interview
	monological	narrative description
		speech formal address
planned	ritual	litany

Biber (1993) proposes a number of additional situational parameters, establishing the catalogue set out in Table 2.

Table 2: *Situational parameters and values, after Biber (1993: 245)*

Parameter	Values	
1. Primary channel	Written/spoken/scripted speech	
2. Format	Published/not published (+ various formats within ‘published’)	
3. Setting	Institutional/other/public/private-personal	
4. Addressee	a) Plurality	Unenumerated/plural/individual/self
	b) Presence (place and time)	Present/absent
	c) Interactiveness	None/little/extensive
	d) Shared knowledge	General/specialised/personal
5. Addressor	a) Demographic variation	Sex, age, occupation, etc.
	b) Acknowledgement	Acknowledged individual/institution
6. Factuality	Factual-informational/intermediate or indeterminate/imaginative	
7. Purposes	Persuade, entertain, edify, inform, instruct, explain, narrate, describe, keep records, reveal self, express attitudes, opinions, or emotions, enhance interpersonal relationship, ...	
8. Topics	...	

A problem not resolved by these classificatory parameters is the fuzzy boundaries between categories, and their nested character. Although all frameworks are aware of the lack of clear boundaries between types of OCEs and their internal heterogeneity, and try to take this into account, e.g. by

dividing speech events into the speech acts they constitute (for instance a lecture can be divided into an opening, a joke, some housekeeping remarks, a question by a student, etc.), it seems impossible to apply such a fine-grained classification to OCEs in a field-based corpus. As a result, all attempts at categorizing OCEs will fail to some extent, however useful an awareness of their possible categorisation is.

A general guideline for the collection of OCEs is that they should respect cultural settings as closely as possible. For instance, the intrusion and potential disruption caused by the presence of observers and recording equipment should be kept as small as possible, only participants normally involved actively or as part of the audience or bystanders should be present, etc. The range of OCEs included will also be influenced by the demographic profile of the collector – some OCEs will only be accessible for collectors of a certain age, sex, linguistic profile, social status, profession, religious affiliation, and so on. When planning the inventory of OCEs to be contained in the corpus, it is advisable to take the profile of the collector into account and to plan ahead and involve other researchers or members of the speech community to record these events.

OCEs are valuable for analysis in a number of linguistic fields. Although they are not structured enough for phonetic and initial phonological analysis (see Ladefoged 2003), the data can be expected to have a high ecological validity regarding the naturalness of high-frequency segmental and suprasegmental phonological, morphological and syntactic phenomena (OCEs in a small corpus are expected to be less good regarding rare phenomena, however). They also allow formulation of first hypotheses on the relationships between form and meaning; nevertheless, since the real-world reference of items used in texts is often obscure, and it is unlikely that their full range of use will occur, the usability of OCEs to determine the extensions and intensions of items is limited. For an investigation of pragmatics, only an extensive documentation of the contexts and felicity conditions of utterances in the metadata and annotations can supply sufficient information. OCEs of particular genres, e.g. narratives, procedural texts, or songs, are of potential interest to a number of disciplines beyond linguistics and can contribute to representation of cultural, historical, ethnobotanical and ethnomusicological knowledge, among others.

In the following sections, I only address only two types of OCEs, monologues and interactive discourse, according to the linguistic data they yield.

4.1 Data resulting from monologues

Data from monologues (or, more accurately, registers with strong monological components, since pure monologues are extremely rare) is commonly included as a documentary appendix to linguistic descriptions, very often presenting stories or historical narratives. For a LDD corpus, researchers should respect and document the particular cultural norms for monologues, which generally also means paying attention to additional participants providing the necessary audience, by-standers, intermediaries, etc. Story telling in a West African context, for instance, requires an audience providing a kind of scripted interaction by uttering an interjection at regular intervals. Without this limited but crucial interactive component, a story would sound sterile and awkward. At the same time, the associated metadata should describe these settings, as they provide important information on cultural aspects, for instance, when and to whom stories are told; who relates and passes on historical and genealogical narratives; how and on which occasions songs, poetry, and ritual genres are performed; what characterises verbal art, etc. If a large number of oral history and/or verbal art texts are to be included in the corpus, it is worthwhile considering relevant research practices (Duranti 2004; Finnegan 1992, 2008; Raleigh Yow 2005; Sherzer 2002).

Written language use falls almost exclusively into the category of monologue. If this modality is used in the speech community, one should aim at its inclusion, even in the case of informal and/or exographic writing traditions, i.e. writing traditions using a different language (Lüpke 2004).

4.2 Data resulting from interactive discourse

Interactive discourse is prototypically represented by conversation. Conversation is often considered a very ‘messy’ genre, and in terms of internal heterogeneity it certainly is very complex. Elaborate frameworks for transcription and analysis of interactive discourse have been developed, especially in conversation and discourse analysis (Georgakopoulou & Goutsos 2004; Schegloff 2007; Schiffrin et al. 2003; Wooffitt 2005). It is useful to study the transcription recommendations of these fields before planning how to include and represent interactive discourse in a corpus, since the interactivity of the genre requires careful consideration of overlaps, turn-takings and how they are signalled, pauses, etc. Software tools like ELAN that allow multi-tiered and time-linked annotation of audio and video recordings facilitate the task of creating, organizing and analysing transcripts. Again, metadata should aim to capture all the relevant parameters on circumstances, settings, participants, etc.

Interactive discourse is often regarded as the quintessence of naturalness. Limits on its usability are similar to those of monologues. Nevertheless, corpora of interactive discourse can be less user-friendly than monological genres, due to the necessarily more complex transcription and annotation. In addition, the quality of recording and transcription is often less than ideal due to the field conditions. Many interactive genres, especially conversation, include a high degree of repetition, parallelisms, and redundancy that should be taken into account when they are used for corpus-based analysis; this can also make them less attractive for other disciplines.

5. Data resulting from staged communicative events

5.1 A hybrid event type

SCEs have no immediate ecological validity: they are communicative events ‘staged’ for the purpose of the research. In that respect, they are very close to elicitations, which are also brought about through direct influence of researchers and motivated by their research interests only. On the other hand, SCEs do not involve direct linguistic influence, since speakers are free to elaborate at will on a verbal prompt like ‘Tell me how you pick mangoes, please’ or to describe a visual prompt in their own words. Lack of linguistic influence in the metalanguage that might cause interference from the prompt crucially distinguishes SCEs from elicitations and makes them close to OCEs in several, but not all, respects. This being said, SCEs, particularly those based on nonverbal stimuli, are frequently aimed at obtaining particular linguistic structures as their response, e.g. the ‘cut and break’ video clips (Majid & Bowerman 2007) which looks at particular aspects of verb semantics.

This video stimulus features a number of cutting and breaking events that either feature two central participants, an Effector⁵ performing the action of cutting and breaking and a Theme, or only one participant, the Theme, undergoing the process. Figure 3 illustrates the stimulus with two stills from videos featuring an Effector and a Theme.

⁵ I follow Van Valin & Wilkins (1996) in distinguishing between Agents (participants that act volitionally and exert control in bringing about the eventuality denoted by the verb) and Effectors (which merely bring about the eventuality denoted by the verb).

Figure 3: Two stills from the ‘cut and break’ clips (Majid & Bowerman 2007)

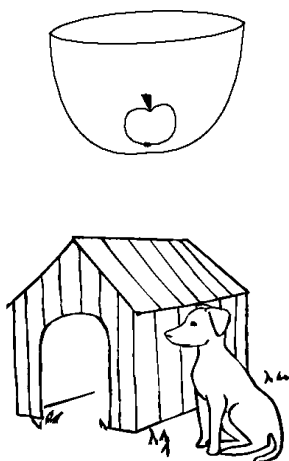


Because of the nature of the depicted scenes, responses to the stimulus are very likely to comprise utterances with roughly parallel syntactic structures (in English likely to be expressed with transitive active clauses like *He chopped the carrots* for externally caused events, and intransitive inchoative clauses like *The twig snapped* for uncaused events). In addition, only a limited number of lexical items will be used, determined by the entities and actions depicted. While these data are more ‘natural’ than elicitations since there is no target from the metalanguage present to influence their linguistic structures, they should not be used for statistical investigations because of the likely preponderance of structural parallelisms. This is even more so if data were collected from more than one consultant. Structural and lexical repetition is expected to increase proportionally with the number of consultants describing a stimulus, and hence only a selection of SCEs should enter a stratified corpus, if at all.

5.2 Considerations for the design and use of stimuli when collecting SCEs

Most stimuli are designed to collect data for specific research questions, and they and the resulting data reflect the underlying research paradigms and theories in which these questions are meaningful. Data collected with a particular stimulus should therefore be analysed preferably in this context, and instructions for the collection and analysis of data, if available, should be adhered to wherever possible. Consider the Topological Relations Picture Series TRPS (Bowerman & Pederson 1993), a widely used picture stimulus containing 71 line drawings. Figure 4 shows two scenes from it.

Figure 4: Two scenes from the Topological Relations Picture Series (Bowerman & Pederson 1993)

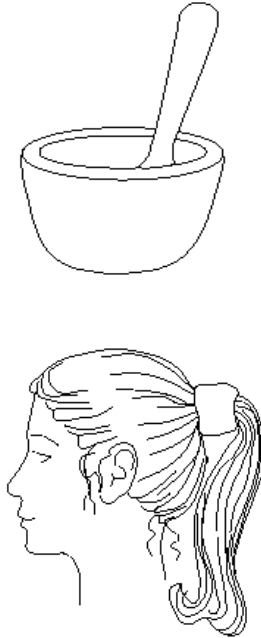


Like most stimuli, TRPS has been developed to collect data on a specific linguistic domain, in this case basic topological relations (as linguistically expressed in the simplest answers to where-questions), including IN⁶, ON, UNDER, OVER, NEAR and AGAINST relations, among others. Levinson & Meira (2003) demonstrate the significance of this stimulus for the

⁶ The use of capital letters signals that these topological concepts are not meant to be exactly those designated by the English words.

development of a semantic typology of topological relations through systematic collection and comparison of cross-linguistic data. The stimulus has been extended to cover the domains of containment and attachment respectively; Figure 5 features two scenes from these areas.

Figure 5: Additions to TRPS from the domains of containment (left) and attachment (right) (Levinson & Enfield 2001)



With TRPS, it is possible to collect data on topological relations that lend themselves to cross-linguistic comparison. However, there is an unavoidable trade-off in the use of stimuli like TRPS which are geared towards the study of specific domains across a wide range of languages: they do not allow a data-driven perspective on the ‘genius’ of a particular language, and they may not have the granularity to uncover all relevant contrasts for a particular linguistic subsystem. Therefore, anyone who wishes to employ stimuli should remind themselves that they come with an inheritance effect – the research question or domain they are originally designed to investigate – and that they cannot substitute for individual-language analyses, e.g. of spatial relation markers. Such individual language analyses would probably explore language-particular functions of the markers as well, for instance their non-spatial uses to locate events in time, or to encode body parts, in addition to the

topological ones. Such an analysis needs to adopt a semasiological perspective, investigating all the functions associated with a given linguistic form, something that cannot be done with stimuli geared at cross-linguistic comparison (which by their very nature are limited to an onomasiological (function-driven) perspective).

In order to make data collected with a stimulus truly comparable, it is also desirable for newly-created stimuli to be accompanied with clear procedures for data collection and coding of results. Only where the original conditions are made explicit and then adhered to as closely as possible are data collected later likely to be ecologically valid (equivalent to OCEs describing similar states of affairs in the real world in spontaneous discourse) and comparable to other data sets.

SCEs yield data that are phonologically, morphologically and syntactically naturalistic, but may present semantic oddities when culturally odd, inappropriate or unusual scenes are depicted. If this is counterbalanced by other data collection methods, however, the data can make important contributions to field-based semantic investigations as well.

5.3 Are stimuli for the collection of SCEs universally applicable?

One criticism of SCEs concerns the lack of universal applicability of visual stimuli, since objects featured in them may be unknown in the field context, or their depiction may violate cultural taboos. The rather problematic claim that consultants lack the capacities for the abstract reasoning necessary to carry out particular stimulus-based tasks is sometimes added to these reservations.⁷ The first two objections are easy to refute: in most cases, it will be possible to negotiate culturally adequate replacement objects with consultants. The apple in Figure 4 may well be substituted with a mango, and the carrots in Figure 3 with cassavas, for example. Care should be taken, however, that any replacement object possesses the linguistically salient properties of the original referent. In the case of the apple in TRPS, for instance, it is important to only replace it with fruit of a circular shape and a growth point on which it can 'sit' and that might be interpreted as its base; otherwise, the results will not be comparable with other research. Regarding the third objection,

⁷ In fact, similar observations hold for problems of translatability when using verbal prompts, such as word lists, questionnaires, etc. These often contain words and constructions without a direct equivalent in the target language or include abstract concepts that may or may not exist in the target language. The conceptual problems are thus comparable for any kind of cross-linguistic investigation comparing concepts, be they linguistic or non-linguistic.

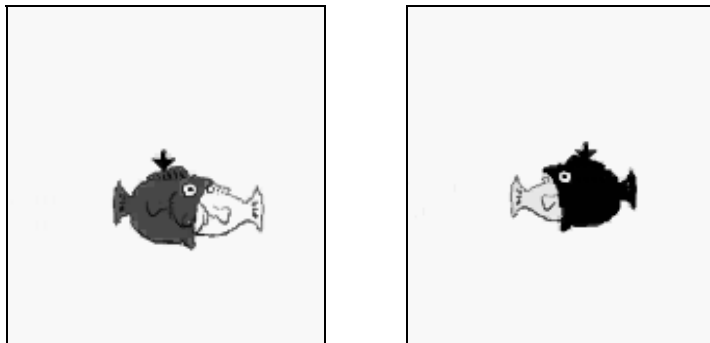
successful use of stimuli depends more on pilot tasks and consultant training than on consultants' inability to understand the task. Unfortunately, many stimuli packages designed for use in the field neglect the training aspect and only contain the core stimuli items. In addition, field researchers are generally unfamiliar with psycholinguistic practice and underestimate the need for consultants to familiarise themselves with the stimuli in question, to counteract repetition through the use of filler items, and to take attention spans and tiredness into account. Many newer stimuli contain training clips with explicit procedures to prepare consultants for the task at hand and result in a less stressful experience for both researchers and consultants. If this is not the case it is advisable to set enough time aside to make consultants and fieldworkers alike relaxed and confident about the task before running it.

Another criticism often voiced about stimuli in field-based research is that they may exhibit structural patterns that are atypical for OCEs in the language and hence might distort analysis – see for instance Foley (2003) on the differences between narratives based on the children's book *Frog, where are you?* and other narratives in Watam. Foley's plea for 'thickest' or varied descriptions can only be repeated here: rather than excluding SCEs, they should be used for specific purposes, and always complemented by OCEs from the widest variety of genres and registers possible. The potential dangers of SCEs leading to biased analyses can be uncovered and minimised by triangulation (through data collection with a wide range of consultants, a balanced variety of methods, situational contexts and communicative event types and, if at all possible, with different researchers).

For all visual stimuli, picture reading preferences based on the directionality of the writing system(s) used by the speech community should be taken into account. There is now compelling psycholinguistic evidence that the orientation of writing systems has relativistic effects on spatial cognition, resulting in different expectations about the thematic roles of participants to the left versus to the right of a picture, and about the flow of action in a sequence of images, etc. (cf. Dobel et al. 2007; Maass & Russo 2003). Since stimuli have been overwhelmingly developed in cultures with Latin-based scripts, there tends to be a bias in the depicted scenes, resulting in Effectors appearing to the left of Themes, Recipients, etc., and sequences of events unfolding from left to right and top to bottom of the page. Where the cognitive impacts of the stimuli themselves are not the focus of the research, the effects are probably negligible. However, researchers should be aware of the dominant writing system used by their consultants and counterbalance its potential influences through training. An example of an animation stimulus that counterbalances possible cognitive effects induced by the writing system is the Fish Film, developed by Tomlin (1995). The Fish Film facilitates the study of thematic structure across languages through 32 clips of two fish that swim onto the screen from the left or right. The left or right fish eats the other

one, and one of the fish (either the Agent or the Theme of the eating event) is highlighted with an arrow cue. The scenes are balanced for direction of entrance and highlighted participant. Figure 6 shows two stills from this stimulus, with different directions of entrance of the Agent but identical cues.

Figure 6: Two stills from the Fish Film (Tomlin 1995)



In the following sections, I introduce SCEs based on verbal prompts before concentrating on those based on nonverbal stimuli, discussing static, dynamic, and interactive stimuli as well as triad tasks and ‘ad hoc’ stimuli.

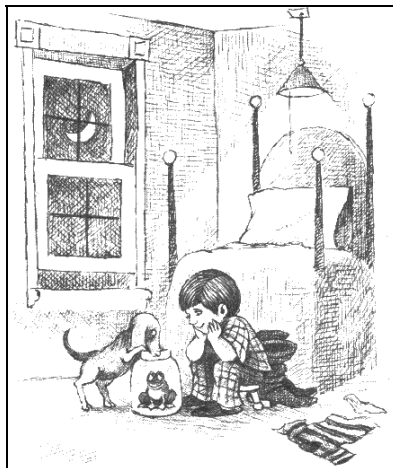
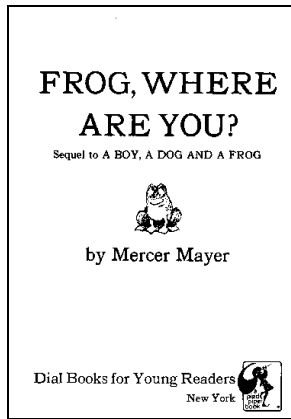
5.4 Data resulting from verbal prompts

Verbal prompts have a long tradition, especially for collecting procedural texts, and will not be addressed in detail here, with one exception: the evoking of contexts for language use in salvage projects on moribund languages. Fieldworkers confronted with these situations, in which the last remaining speakers of the dying language often lack any contexts for speaking their language, require special techniques to activate long-term memory. A number of researchers working in these contexts report using verbal and synaesthetic prompts to facilitate memory. They verbally evoke situations associated with use of the language, e.g. harvest festivals, religious ceremonies, occasions for story telling, etc. in order to activate forgotten associations. Where possible, this task is accompanied by synaesthetic prompts like exposing speakers to artefacts, locations, photos, or smells to spur memory even further – these might also be called nonverbal stimuli. Since these situations constitute quite special cases, however, they deserve a separate mention here.

5.5 Data resulting from static stimuli

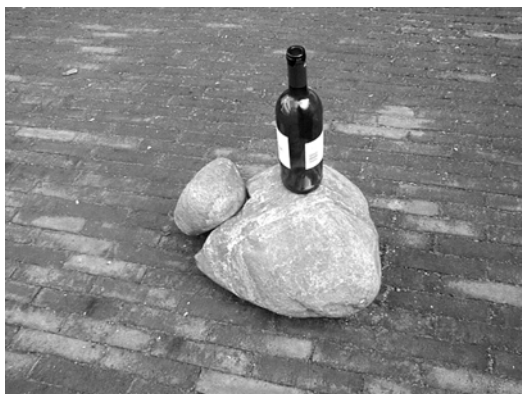
Static stimuli probably have the longest history of use of all nonverbal stimuli. Examples include the drawings in the groundbreaking collection of (ethno)linguistic stimuli and questionnaires in Bosquiaux & Thomas (1971) to collect terms for body parts, flora, fauna, etc. The previously mentioned children's book *Frog, where are you?* (Figure 7) has been a widely-used stimulus in cross-linguistic research on acquisition of narrative structure (Berman & Slobin 1994).

*Figure 7: Front cover and scene from the children's book *Frog, where are you?* (Berman & Slobin 1994)*



While older static stimuli mainly consist of line drawings, the use of photographs has been gaining ground due to the ease of digital photography; an example of a photo stimulus is the Picture Series for Positional Verbs (henceforth PSPV, Ameka et al. 1999) devised to uncover the extensions of positional verbs (e.g. English *sit*, *stand* and *lie*) in the description of single and multiple located objects across languages (see Figure 8).

Figure 8: Two scenes from the Picture Series for Positional Verbs (Ameka et al. 1999)



5.6 Data resulting from dynamic stimuli

Dynamic stimuli fall into two groups: video clips and animations. Animations are mainly used for abstract or unrealistic scenes or in stimuli where parameters need to be varied in a detailed and controlled manner unachievable by human actors. In LDD research, video stimuli predominate since digital video has become affordable and user-friendly. The earliest known video stimulus to my knowledge is Wallace Chafe's Pear Film made in 1975, devised to collect narratives through recounting the content of a 6 minute film according to a standard procedure (see Chafe 1980). Figure 9 shows two stills from the Pear Film; data from 17 languages have been collected using it.

Figure 9: Stills from the video stimulus The Pear Film (Chafe 1980)



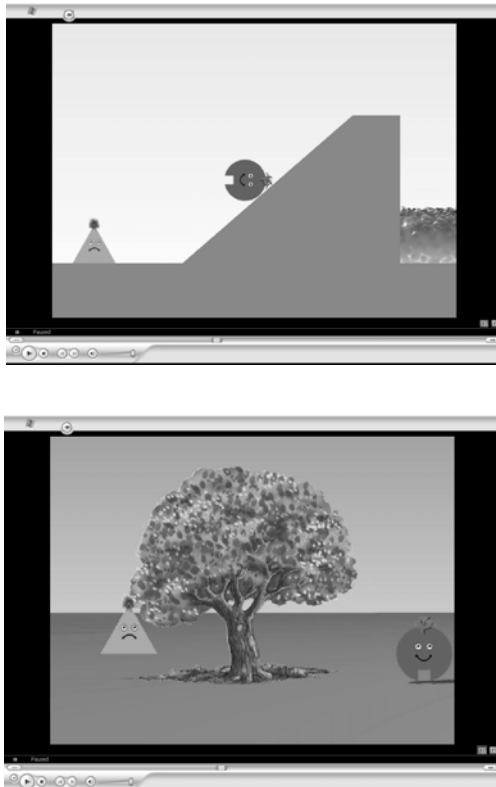
Widely-used examples of video stimuli include the Cut and Break (Majid & Bowerman 2007) and Staged Event clips (van Staden et al. 2001), and a number of other video stimuli developed by the Max Planck Institute for Psycholinguistics, Nijmegen. Since video stimuli are able to show events unfolding over time, they can cover domains not easily depicted in static stimuli. Take the Caused Positions stimulus (Hellwig & Lüpke 2001), for instance. This video stimulus is inspired by the static stimulus PSPV, which investigates the use of positional verbs for simple (stative) positions (cf. English *The cassavas are lying on the table*). The dynamic nature of video allows extension of this investigation to caused and uncaused position changes (cf. English *She put the cassavas on the table*). Figure 10 shows two stills from the Caused Positions stimulus.

Figure 10: Two scenes from the video stimulus Caused Positions (Hellwig & Lüpke 2001)



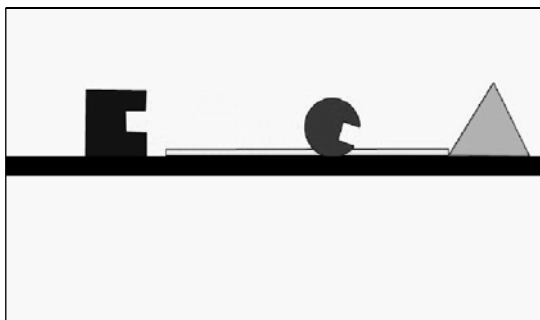
Animation stimuli are also widely used. When created with software, they allow controlled manipulation of visual parameters that acted video stimuli cannot easily achieve – e.g. the speed of actions can be adjusted even in milliseconds, backgrounds for scenes can be changed with a mouse click, the directions of entrance and exit of the protagonists can be reversed, etc. However, the software necessary to create such sophisticated animations have not been widely used in field-based linguistic research, mainly because the applications are expensive and highly complex to master, and exploitation of all their features requires a considerable investment of time. Therefore, animation stimuli designed and used by fieldworkers tend to be simpler in visual design. This does not reduce their usability for LDD, however. An example is *Tomato Man* (Özyürek et al. 2001), created to investigate the relationship between language and gesture in spatial language and containing two practice and 12 test items (see Figure 11).

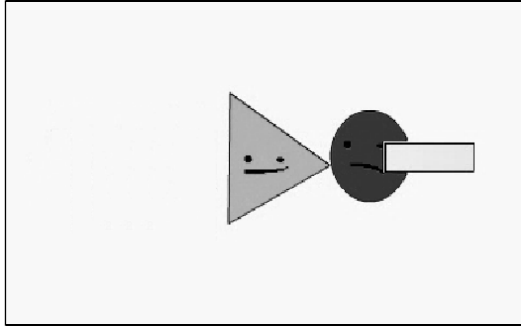
Figure 11: Stills from the animation stimulus Tomato Man (Özyürek et al. 2001)



The Tomato Man animations explicitly invite participants to interpret the red circle and the green triangle as human participants – as Tomato Man and Green Man respectively. Other animation stimuli do not address the danger of ambiguous interpretations. Very often consultants attribute properties of animate, often human, Effectors to moving images of inanimate objects. This interpretation, labelled ‘anthropomorphism’, can have huge effects in the choice of lexical items and constructions used to describe the stimuli, e.g. the actions of animate Effectors are more often described with Effector-prominent constructions (active clauses, caused-motion verbs, etc.), and events featuring them tend to be described as purposeful actions over which they have control (cf. English *Tomato Man pushed Green Man down the hill* versus *The red ball touched the green triangle and it went down the hill*). In contrast to Tomato Man, other animation stimuli lack clear instructions on the interpretation of the moving entities featured in them, e.g. the ECOM clips (Bohnenmeyer & Caelen 1999), created to study the cross-linguistic expression of complex motion and state-change events (Bohnenmeyer et al. 2007). The ECOM clips comprise 74 scenes, grouped into 14 sets, and also come with a set of training items. While a subset of the clips has animacy clues – faces drawn onto the geometrical figures – it is less clear how the moving entities without these clues are to be interpreted by the consultants. Compare the two stills in Figure 12. The geometrical objects in the left-hand image still look very similar to those in the right-hand one, setting aside the animacy clues, and in addition they have indentations allowing them to ‘pick up’ other objects, such as sticks. This creates a danger of unwanted anthropomorphic versus inanimate interpretations, especially for consultants who have already been exposed to explicitly anthropomorphizing stimuli such as Tomato Man. If no clear instructions come with the stimulus, fieldworkers should remind themselves of the dangers of such ambiguous scenes and try to achieve consistency in interpretation through their instructions.

Figure 12: Two stills from ECOM clips (Bohnenmeyer & Caelen 1999)

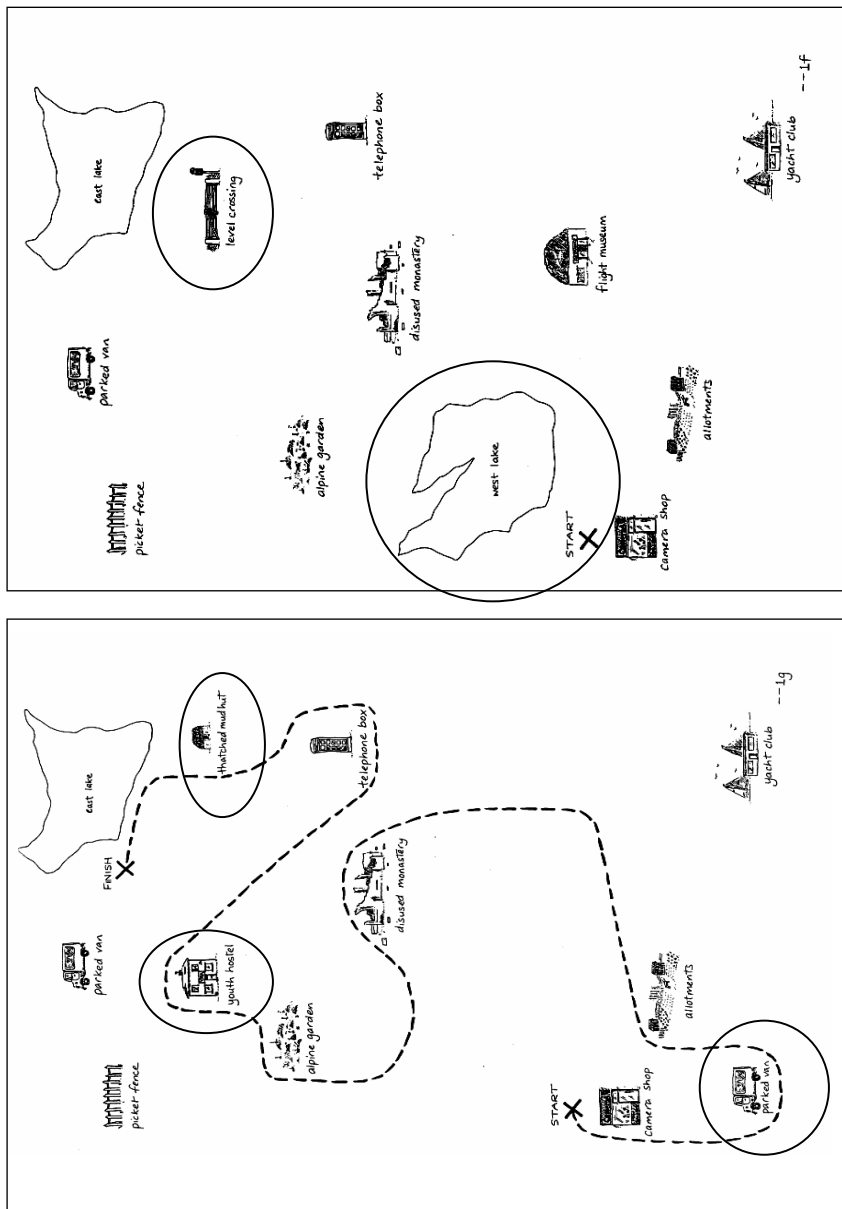




5.7 Data resulting from interactive stimuli

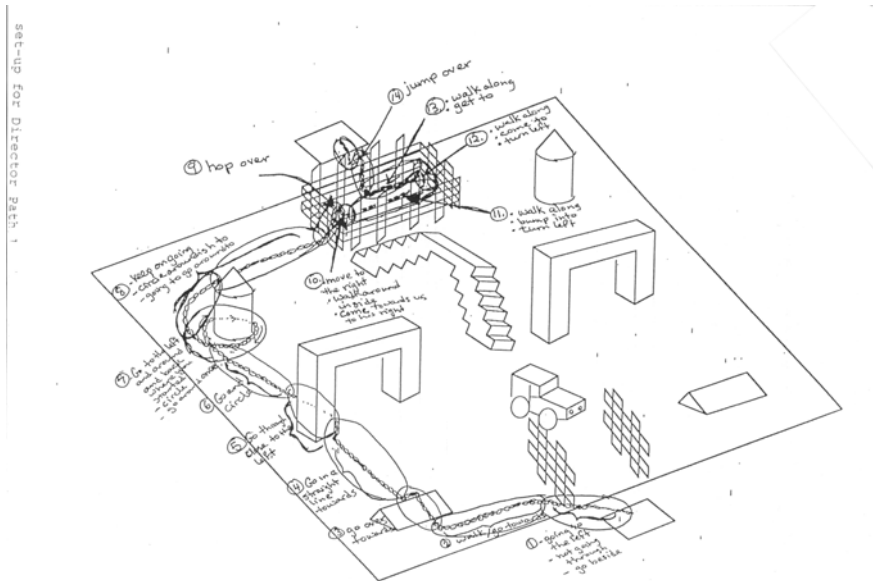
Interactive stimuli to engage two or more consultants in an interaction that is controlled by the nature of the stimulus (as against a consultant responding to a set of stimuli, aided by the researcher). With this stimulus type, very often the true purpose of the interaction (the particular linguistic research goal) is masked and inaccessible to the consultants. A well-known instance is the Map Task developed by the Human Communication Research Centre Glasgow and Edinburgh (Anderson et al. 1991), where two consultants are given two maps with – crucially – slightly different landmarks on them. One of the consultants, the instruction giver, has a route passing a number of the landmarks on their map and describes the route to the second consultant, the instruction follower. The follower’s task is to successfully recreate the route on their map. Since the two maps differ, the two participants discover that they need to engage in an explicit verbal exchange. The Map Task has been used for a wide range of linguistic research– from phonetics to sociolinguistics. It can be easily adapted to suit individual researchers’ needs, e.g. to ensure utterances containing words with certain phonetic properties (featured as landmarks on the maps), etc. Figure 13 shows sample maps.

Figure 13: An instruction giver's map (left) and an instruction follower's map (right), with the differing landmarks circled (Anderson et al. 1991)



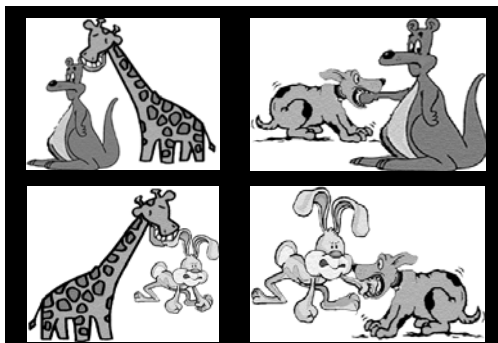
A similar stimulus created to elicit spatial descriptions is the Table Top Route Descriptions task (CARG 1993a). Here Lego blocks are used to build two identical landscapes next to each other, separated by a barrier. The instructor is placed in front of one of the landscapes, in which a route is marked with a chain. The instructor's task is to describe the route to the follower, so that (s)he can successfully recreate it.

Figure 14: Sketch for a Table Top Route Description task (CARG 1993a)



Other interactive stimuli include puzzle tasks and matching games. Figure 15 depicts a puzzle task developed to compare the acquisition of external and internal possession (cf. English *The giraffe bites the kangaroo on the ear* versus *The giraffe bites the kangaroo's ear*) in German and Japanese. In the task, children are given a mould with holes to be filled with puzzle pieces. In order to obtain the matching puzzle piece, they need to describe it to the researcher.

Figure 15: A child doing the puzzle task (left) and scenes featured on puzzle pieces (right) Eisenbeiss et al. 2009)



Card matching card games follow the same logic as map tasks and puzzle tasks: two consultants are given two sets of identical cards containing scenes that are similar, but that vary according to a number of salient parameters. They cannot see each others' cards, and hence need to engage in a dialogue in order to match the cards. Figure 16 shows cards from the game 'Men and Tree' (CARG 1993b), which was devised to establish the frame of reference used by the speakers for spatial orientation – depending on whether an absolute (e.g. north vs. south), intrinsic (e.g. back vs. front) or relative (e.g. left vs. right) frame of reference is employed, different cards would match with each other (see Levinson 2003 for extensive discussion).

Figure 16: Three cards from the matching game *Men and Tree* (CARG 1993b)

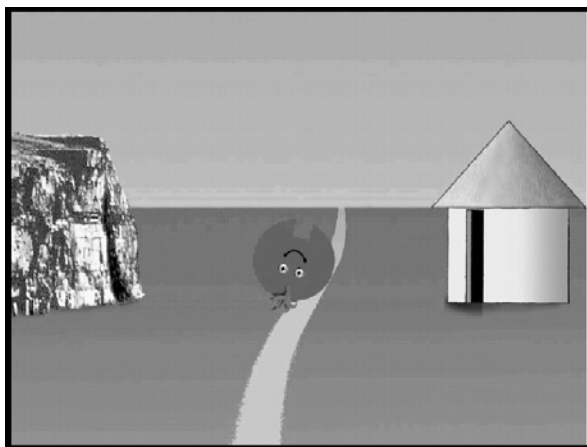


For interactive stimuli, even more so than for ‘responsive’ ones, researcher reflection on culture-specific norms of interaction is required before using them in the field. The linguistic nature of the tasks is transparent only to the researcher. From the consultants’ perspective, the task constitutes a competitive social interaction. Therefore, it is highly advisable to investigate how games of this kind are carried out in the culture under investigation, if a comparable genre exists at all, and to pay attention to preferable pairings for instructors and matchers in terms of differences in age, gender, social group, etc. In this way it will be possible to avoid conflicts between consultants that might potentially undermine the research project and have a negative impact on the long-term relationships between researchers and consultants.

5.8 Data resulting from triad tasks

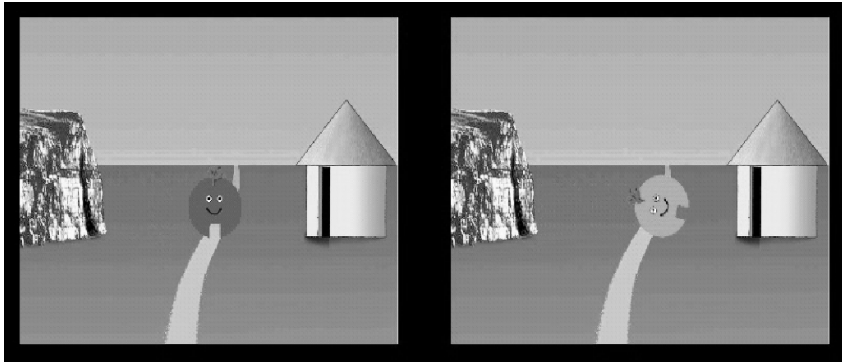
Triad tasks belong to the most psycholinguistically-minded field-based stimuli tasks. Rather than eliciting linguistic descriptions (for which they can also be used), they aim to uncover cognitive principles that potentially underlie differences in grammatical organisation between languages. For example, the Event Triads (Bohnenmeyer et al. 2001) investigates Whorfian effects in the perception of manner versus path of motion depending on how the consultant’s language encodes these components of motion events are explored. To this end, consultants are shown three video clips. The first video clip shows the target scene; in Figure 17 it is a red anthropomorphised circle rolling from a rock to a house.

Figure 17: Still from a target clip from Event Triads (Bohnenmeyer et al. 2001)



The following two clips, shown simultaneously on a split screen, present two scenes that differ crucially from the target clip, either in manner of motion (e.g. rolling versus spinning) or in other parameters (e.g. identity of moving figure, path of motion, etc.). In Figure 18, the circle on the left screen hops from the rock to the house, whereas the circle on the right screen rolls from the house to the rock. Consultants have to point out the clip that most resembles the target clip (see Bohnemeyer et al. (2006) for the results of a cross-linguistic study using this stimulus and a critical review of its methodology).

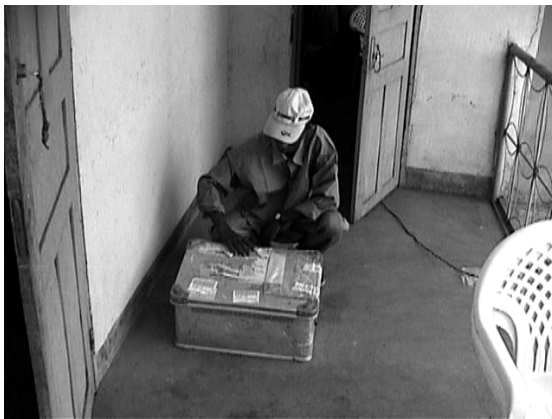
Figure 18: Still showing the split screen clips following the target clip in Figure 17 from Event Triads (Bohnemeyer et al. 2001)



5.9 Data resulting from ‘ad hoc’ stimuli

A final type of stimuli defies easy classification – under the label ‘ad hoc’ stimuli I include those that researchers can develop more or less spontaneously in the field to assist data collection. Such stimuli could be photographs of flora and fauna to aid identification and establishment of taxonomies, or a series of stills showing sequences of an event, like those in Figure 19.

Figure 19: Sequences of stills showing closing a box (Lüpke 2005b)



Photographs like these were invaluable for my investigation of lexical aspect in Jalonke (Lüpke 2005b), because I could segment events into different phases and study the compatibility of the verbs describing them with different tense/aspect markers.

Another example of an ad hoc stimulus that proved very useful in the field is what I have called ‘action descriptions’ in Lüpke (2005a) – videos of everyday events, such as working in the fields, doing laundry, baptizing a child, etc., taken at my field site. I showed these films to consultants and obtained narratives that in some respects are very close to procedural texts. The advantage of action descriptions based on video compared to responses to ‘How do you ...?’ questions lies in the level of granularity. When invited to explain how they, for instance, cultivate cassava, consultants would often reply rather summarily with one or two sentences. When describing an equivalent event shown in a video, they described it without reverting to scripted descriptions but told what they saw as the actions unfolded scene after scene. The absence of a narrative goal yielded very rich and detailed description offering a wealth of vocabulary for the actions and objects featured in the films.

Figure 20: Two stills from action descriptions filmed in Guinea (Lüpke 2005a)





6. Data resulting from elicitation

Elicited data have very low ecological validity – they come into existence under the control of the researcher and are entirely motivated by their research questions. Yet in the history of language description, a considerable proportion of descriptive statements, particularly those supporting linguistic theorising, were arrived at through elicitation. Elicitation remains a very controversial method, with a body of specialised literature dedicated to dissecting its benefits and limits. For language documentation, the role of elicitation and the place of the data resulting from it in a documentary corpus have not yet been discussed in detail. It is not generally contested that elicited evidence is necessary to complement other kinds of corpus data, particularly regarding negative evidence (see 2.6 above). Nevertheless, caution regarding its value is in order – sceptics like Himmelmann (2006a: 23) argue that:

[...] with regard to the usual way of obtaining negative evidence (i.e. asking one or two speakers whether examples *x*, *y*, *z* are “okay”), it is doubtful whether this really makes a difference in quality compared to evidence provided by the fact that the structure in question is not attested in a large corpus.

Dimmendaal (2001) and Mithun (2001) also question the value of elicitation, particularly when it consists of grammaticality judgments, as does Schütze (1996, 2005) for English, where, at least partially, more sophisticated techniques than those sketched by Himmelmann have been used. For LDD,

some recommendations are introduced in the following sections on the elicitation of translation equivalents and acceptability judgments – however in view of the complexity of the issue, I make no claim to completeness.⁸

6.1 Data resulting from the elicitation of translation equivalents

Almost automatically, most field linguists collect word lists as the first elicitation conducted at the beginning of a new field-based LDD project. Word lists are often recommended in field methods classes as an easy way to start gathering lexical data on an unknown language, and different regional traditions of descriptive linguistics have their own versions of widely-used word lists, such as the Swadesh lexicostatistical list. Yet eliciting word lists from consultants is a very tedious means of data collection, and is fraught with unwelcome dangers, such as lack of a direct translation equivalent, misunderstandings and lack of proficiency in the common language (on the part of the researcher and/or the consultants), interference from the shared language resulting in calques, etc. (see Bradley 2007). Rather than discarding word lists as an instrument altogether, however, I suggest viewing them as the result of data collection rather than as an elicitation tool. Other methods bypassing them have been suggested and yield more reliable data; Mosel (2004, 2006) advocates collecting lexical data by starting with semantic fields, and either grouping word list items or asking consultants for names of, for instance, fruits and vegetables, components of a house, and the like. Stimuli-supported methods of obtaining vocabulary items such as showing pictures and asking for the names of the objects in them are also a possible way to circumvent problems with word lists. Ladefoged (2003) discusses different methods to obtain word lists containing a minimal pair needed for the initial phonological analysis of a language: compile the words from a short text, complement them with nonsense words that exhibit the desired contrast and ask if they exist in the language, and seek out possible rhymes and alliterations for attested words.

In later stages of fieldwork it is very likely that concordances produced from corpus data will yield additional information, e.g. on collocations, contexts of use and morphosyntactic status of lexical items. In addition,

⁸ In particular, monolingual elicitation is not treated. Everett (2001) gives an account of the challenges and benefits of not using a lingua franca in the field. However, I do believe that in most fieldwork situations, at least initially and for a number of tasks, a common language is both indispensable and useful, and therefore do not discuss monolingual fieldwork here.

monolingual explanations and encyclopaedic information will allow for deeper insight into intensions and extensions.

Another area is translation of isolated sentences from a shared language into the target language, often by using questionnaires for a particular domain (e.g. causatives or valency-changing operations). Questionnaires are a tool widely used by typologists, but most (e.g. the influential *Lingua Descriptive Studies* questionnaire of Comrie & Smith 1977) are devised as lists of target structures in the common language, rather than as instruments for data collection. Yet often field linguists ask for translation equivalents of sentences given as examples in questionnaires, a procedure that has numerous undesirable side effects, as addressed in detail by Matthewson (2004). While she argues for the use of translation equivalents as an elicitation method in LDD, she also discusses the various levels at which that method can fail.

In order to avoid undesirable misunderstandings, for instance resulting from the structural ambiguity of an NP, uncertainty of interpretation of ungrammatical structures, etc., Matthewson (2004: 388) recommends adhering to the following guidelines for the collection of translation equivalents:

1. ask for translations of complete sentences only;
2. try to make the source string a grammatical sentence;
3. assume that the result string is a grammatical sentence.

She warns against inviting consultants to analyse sentences, against regarding translations as a result rather than as a clue, and against taking seriously apparent information about the felicity conditions of translated sentences. Regarding the often necessary discourse context required to avoid ambiguity or vagueness, she further suggests providing the appropriate discourse context before the sentence is translated, and using non-verbal cues (see 5 above) if appropriate.

6.2 Data resulting from judgment tasks

There is no unanimity among linguists about what can legitimately be called a ‘linguistic judgment’, and even less about who – the speaker, the linguist, the native speaker linguist... – is qualified to make it. Matthewson (2004) recognises only three types of judgments: grammaticality judgments, truth value judgments and felicity judgments. Schütze (1996) additionally admits ambiguity judgments, which Matthewson relegates to linguistic analysis, along with judgments about the correctness of paraphrases, another often-mentioned judgment type. Disputes about the appropriate judgment of English sentences like *Spiro conjectures Ex-Lax* have fuelled linguistic

debates (Harris 1993). Yet judgments, particularly on ungrammaticality, are ubiquitous in linguistic descriptions, where they are used to make strong assertions about language structure, and therefore it is worthwhile to investigate what good practices could look like for obtaining them under field conditions. In the absence of explicit and universally recognised representations for (degrees of) (un)grammaticality, it is important to define and document symbols and scales used for judgment ratings. In addition, a number of factors have been shown to have an influence on the outcome of judgment tasks. Schütze (1996) offers a useful classification of these factors, dividing them into task-related, subject-related and stimulus factors. Here I list the factors that I consider relevant for field-based research and complement Schütze's recommendations with my own suggestions for an LDD context, where appropriate.

Task-related factors which have been shown in experimental research to have an impact on the outcome of judgment tasks include the following:

1. the nature of the instructions. Without clear instructions explaining exactly how 'grammatical' or 'acceptable' are to be understood, judgment tasks have been shown to yield erratic responses. Schütze proposes working with consultants prior to the task to arrive at an explanation that is accurate, yet clear, and then presenting the agreed upon example sentences and their ratings to the participants in the task and explaining to them the grounds on which these were chosen;
2. order of presentation. The first sentence presented in a judgment task is consistently rated lower than its successors. Schütze recommends randomising or counterbalancing the order of presentation and having training tasks (see also 5.3) that precede the test items;
3. repetition of task items. The rating of sentences can change when they are repeated. The psycholinguistic evidence is not entirely conclusive, so from an LDD perspective I would suggest repeating tasks after certain intervals if at all possible to check if ratings have changed, and if so, investigate the reasons.

A number of considerations are relevant to the choice of consultants for judgment tasks. Here, Schütze lists the following factors, alongside

individual differences between subjects: field-dependence⁹, history of handedness, and experience and training. Field-dependence and history of handedness in a consultant's family are factors difficult to control during fieldwork; I consequently limit myself to the following **subject-related factors**:

1. training in, experience with, and exposure to judgment tasks. These factors can result in dramatic differences across subjects, but the available evidence is inconclusive regarding whose judgment is more stringent or reliable. Therefore, my recommendation for LDD is to be explicit in collecting and documenting information on consultants' personal profiles, including experience in metalinguistic tasks, or, if possible, apply a set of criteria in order to form a homogenous pool of consultants for judgment tasks.
2. literacy and educational background. Literacy, more specifically literacy acquired in formal settings, has been attributed to have a large impact on a number of tasks involving abstract reasoning (e.g. Scribner & Cole 1981). In order to eliminate the unwanted influence of literacy and education, I suggest to selecting consultants with similar educational backgrounds and to document this background.

Finally, some attention should be given to possible effects of the stimulus on the outcome of judgment tasks. Schütze gives the following list of **stimulus-related factors**:

1. context. The role of context for judgment sentences cannot be underestimated. Many studies are flawed because no context is provided, which simply means that the interpretation is beyond the researchers' control. Yet, it is not sufficient just to provide some context, because different types of contexts may have different impacts on the rating of the sentence. I suggest that a 'neutral' or 'stereotypical context' may be the best for judgment tasks not explicitly testing the limits of grammaticality of a given construction;

⁹ In brief, field-dependent persons experience the world globally, whereas field-independent persons differentiate their experiences into compounds. Psychological tests such as picking out geometric figures that are embedded into bigger ones allow us to distinguish between field-dependent subjects, who fare worse in this test, from field-independent ones who do better. While there is evidence that field-dependence has an influence on linguistic judgment tasks, no clear patterns emerge.

2. meaning. It is questionable to what extent it is possible to rate grammaticality of a structure independently of the meaning conveyed through it. I therefore recommend investigating the two features separately, if possible, i.e. grammaticality judgment tests should not contain nonsensical sentences like *Colourless green ideas sleep furiously*, and semantic judgment tests should not contain ungrammatical strings like *One browns cat*;
3. parsability. There is evidence that the length and complexity of sentences affect their rating. Therefore, sentences in a task should be balanced according to these parameters;
4. frequency. The frequencies in which structures occur in a corpus are closely correlated with their acceptability in judgment tasks. If the relevant data are available, frequency information should therefore accompany the ratings of constructions;
5. lexical content. Individual words in sentences have been demonstrated to affect their rating. If lexical items have imagery content, sentences are rated higher;
6. morphology and spelling. The transparency of morphology, and the frequency of a given spelling for written stimulus sentences have been reported to have an influence on ratings. Therefore, these factors should be kept constant in sentences to be compared.

In LDD, a number of other factors need to be taken into account. These particularly relate to the linguistic variation and absence of standardisation encountered in most LDD contexts and will be discussed in the conclusion, since they are also relevant for other data types.

7. Conclusion

This paper has argued for a more reflexive approach to data collection methods than commonly practiced in language documentation and description. In particular, it has argued for a greater awareness of different methods of data collection and the kinds of data they yield. If we want to arrive at descriptive statements that reliably reveal the structure of the languages we study, we need to find robust and testable methods for arriving at them. Driven by considerations of variation and introspection, Labov (1975: 40) coined the following widely-cited principles for arriving at valid linguistic facts:

- I. *Consensus Principle*: if there is no reason to think otherwise, assume that the judgments of any native speaker are characteristic of all speakers of the language.
- II. *Experimenter Principle*: if there is any disagreement on introspective judgments, the judgments of those who are familiar with the theoretical issues may not be counted as evidence.
- III. *Clear Case Principle*: disputed judgments should be shown to include at least one consistent pattern in the speech community or be abandoned. If differing judgments are said to represent different dialects, enough investigation of each dialect should be carried out to show that each judgment is a clear case in that dialect.
- IV. *Principle of Validity*: when the use of language is shown to be more consistent than introspective judgments, a valid description of the language will agree with that use rather than introspections.

From an LDD perspective, and guided by portability issues (Bird & Simons 2003), I would like to add the following five principles to Labov's:

- V. *Principle of Explicitness*. Analytical choices and decisions should be made explicit, i.e. the reasons for selecting a particular data collection method, including or excluding a particular set of data, and working with a specific (group of) consultant(s) should be documented in metadata descriptions and annotations of primary data;
- VI. *Principle of Transparency*. Abbreviations, symbols, labels, meanings of tiers used in transcriptions and annotations, numeric variables in spreadsheets, etc., should be explained in metadata and annotations of primary data;
- VII. *Principle of Salience*. For the analysis of a particular research question, the most salient method for collection and analysis should be selected. For instance, descriptions of visual scenes rather than translation equivalents should serve as the basis for the analysis of spatial language;
- VIII. *Principle of Triangulation*. Wherever possible, analysis should be verified through triangulation, i.e. through different methods of data collection, data from more than one consultant, different types of analysis, and comparison of data with those collected by other researchers, etc., whenever possible.
- IX. *Principle of Longevity*. Efforts should be made to make data valid beyond the scope of the particular research by not just seeking the data necessary to answer specific research questions or relating to one particular area of language use. So, for instance, when collecting data on the encoding of topological relation, researchers should not limit themselves to stimulus-based data collected with TRPS but complement these data with OCEs containing spatial descriptions, etc.

The issues and recommendations raised in this paper, join a growing body of suggestions for good practices in LDD (see Austin 2006; Chelliah 2001; and Woodbury 2003, to name but a few others). I hope that this constitutes a start in the development of reflective methodological discussion and the emergence of a canon of research practices following the example of the social sciences, and that fieldworkers, theoreticians of LDD and neighbouring disciplines, along with members of speech communities, will contribute to driving forward the methodological agendas in this important field.

References

- Ameka, Felix, Carlien de Witte & David Wilkins. 1999. Picture series for positional verbs: Eliciting the verbal component in locative descriptions. In David Wilkins (ed.), *'Manual' for the 1999 field season*, 48-56. Nijmegen: Max Planck Institute for Psycholinguistics, Language and Cognition Group.
- Ameka, Felix K., Alan Dench & Nicholas Evans (eds.), 2006. *Catching language: The standing challenge of grammar-writing* (Trends in Linguistics: Studies and Monographs 167). Berlin: Mouton de Gruyter.
- Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson & Regina Weinert. 1991. The HCRC map task corpus. *Language and Speech* 34(4), 351-366.
- Austin, Peter K. 2006. Data and language documentation. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics: Studies and Monographs 178), 87-112. Berlin: Mouton de Gruyter.
- Austin, Peter K. & Lenore A. Grenoble. 2007. Current trends in language documentation. In Peter K. Austin (ed.), *Language documentation and description*, vol. 4, 12-25. London: SOAS.
- Barwick, Linda. 2005. A musicologist's wishlist: Some issues, practices and practicalities in musical aspects of language documentation. In Peter K. Austin (ed.), *Language documentation and description*, vol. 3, 53-62. London: SOAS.
- Bergqvist, Henrik. 2007. The role of metadata for translation and pragmatics in language documentation. In Peter K. Austin (ed.), *Language documentation and description*, vol. 4, 163-173. London: SOAS.
- Berman, Ruth A. & Dan Isaac Slobin. 1994. *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bernard, H. Russell. 2002. *Research methods in anthropology: Qualitative and quantitative approaches*, 3rd edn. Walnut Creek, CA: AltaMira Press.

- Biber, Douglas. 1990. Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing* 5(4), 257-269.
- Biber, Douglas. 1993. Representativeness in corpus design. *Literary and Linguistic Computing* 8(4), 243-257.
- Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Bird, Steven & Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language* 79(3), 557-582.
- Bohnenmeyer, Jürgen & Martijn Caelen. 1999. The ECOM clips: A stimulus for the linguistic coding of event complexity. In David Wilkins (ed.), *'Manual' for the 1999 field season*, 74-86. Nijmegen: Max Planck Institute for Psycholinguistics, Language and Cognition Group.
- Bohnenmeyer, Jürgen, Sonja Eisenbeiss & Bhuvana Narasimhan. 2001. Event triads. In Steve Levinson & Nick Enfield (eds.), *'Manual' for the field season 2001*, 100-114. Nijmegen: Max Planck Institute for Psycholinguistics, Language and Cognition Group.
- Jürgen Bohnemeyer, Sonja Eisenbeiss & Bhuvana Narasimhan. 2006. Ways to go: Methodological considerations in Whorfian studies on motion events. *Essex Research Reports in Linguistics* 50, 1-20. <http://www.essex.ac.uk/linguistics/publications/erri/> (1 February 2010).
- Bohnenmeyer, Jürgen, Nicholas J. Enfield, James Essegbey, Iraide Ibarretxe-Antuñano, Sotaro Kita, Friederike Lüpke & Felix K. Ameka. 2007. Principles of event segmentation in language: The case of motion events. *Language* 83(3), 495-532.
- Bouquiaux, Luc & Jaqueline M. C. Thomas. 1971. *Enquête et description des langues à tradition orale*. Paris: SELAF (Societe d'Etudes Linguistiques et Anthropologiques de France).
- Bowerman, Melissa & Eric Pederson. 1993. Topological relations picture series. In Eve Danziger & Deborah Hill (eds.), *Manual for the space stimuli kit 1.2*, 40-50. Nijmegen: Max Planck Institute for Psycholinguistics, Cognitive Anthropology Research Group.
- Bradley, David. 2007. What elicitation misses: Dominant languages, dominant semantics. In Peter K. Austin (ed.), *Language documentation and description*, Vol. 4, 136-144. London: SOAS.
- CARG (Cognitive Anthropology Research Group). 1993b. Men and Tree. In Eve Danziger & Deborah Hill (eds.), *Manual for the space stimuli kit 1.2*, 11-13. Nijmegen: Max Planck Institute for Psycholinguistics, Cognitive Anthropology Research Group.
- CARG (Cognitive Anthropology Research Group). 1993a. 'Route description' elicitation. In Eve Danziger (ed.), *Cognition and space kit 1.0*, 15-28. Nijmegen: Max Planck Institute for Psycholinguistics, Cognitive Anthropology Research Group.
- Chafe, Wallace L. (ed.). 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (Advances in Discourse Processes 3). Norwood, NJ: Ablex Publishing Corporation.

- Chelliah, Shobhana L. 2001. The role of text collection and elicitation in linguistic fieldwork. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, 152-165. Cambridge: Cambridge University Press.
- Comrie, Bernard & Norval Smith. 1977. Lingua descriptive studies: Questionnaire. *Lingua* 42(1), 1-72.
- Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Basil Blackwell.
- Dimmendaal, Gerrit J. 2001. Places and people: Field sites and informants. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, 55-75. Cambridge: Cambridge University Press.
- Dobel, Christian, Gil Diesendruck & Jens Bölte. 2007. How writing system and age influence spatial representations of actions: A developmental, cross-linguistic study. *Psychological Science* 18(6), 487-491.
- Duranti, Alessandro (ed.). 2004. *A companion to linguistic anthropology* (Blackwell Companions to Anthropology 1). Oxford: Blackwell.
- Dwyer, Arieane M. 2006. Ethics and practicalities of cooperative fieldwork and analysis. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics: Studies and Monographs 178), 31-66. Berlin: Mouton de Gruyter.
- Eisenbeiss, Sonja. 2005. Documenting child language. In Peter K. Austin (ed.), *Language documentation and description*, vol. 3, 106-140. London: SOAS.
- Eisenbeiß, Sonja, Ayumi Matsuo & Ingrid Sonnenstuhl (2009). Learning to encode possession. In William B. McGregor (ed.), *The expression of possession* (The Expression of Cognitive Categories 2), 143-211. Berlin: Mouton de Gruyter.
- Evans, Nick & Hans-Jürgen Sasse. 2007. Searching for meaning in the Library of Babel: Field semantics and problems of digital archiving. In Peter K. Austin (ed.), *Language documentation and description*, vol. 4, 58-99. London: SOAS.
- Everett, Daniel L. 2001. Monolingual field research. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, 166-188. Cambridge: Cambridge University Press.
- Finnegan, Ruth. 1992. *Oral traditions and the verbal arts: A guide to research practices* (ASA Research Methods in Social Anthropology). London: Routledge.
- Finnegan, Ruth. 2008. Data – but data from what? In Peter K. Austin (ed.), *Language documentation and description*, vol. 5, 13-28. London: SOAS.
- Florey, Margaret. 2004. Countering purism: Confronting the emergence of new varieties in a training program for community language workers. In Peter K. Austin (ed.), *Language documentation and description*, vol. 2, 9-27. London: SOAS.

- Foley, William A. 2003. Genre, register and language documentation in literate and preliterate communities. In Peter K. Austin (ed.), *Language documentation and description*, vol. 1, 85-98. London: SOAS.
- Foley, William A. 2005. Personhood and linguistic identity, purism and variation. In Peter K. Austin (ed.), *Language documentation and description*, vol. 3, 157-180. London: SOAS.
- Franchetto, Bruna. 2006. Ethnography in language documentation. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics: Studies and Monographs 178), 183-211. Berlin: Mouton de Gruyter.
- Georgakopoulou, Alexandra & Dionysis Goutsos. 2004. *Discourse analysis. An introduction*, 2nd edn. Edinburgh: Edinburgh University Press.
- Grenoble, Lenore A. 2007. The importance and challenges of documenting pragmatics. In Peter K. Austin (ed.), *Language documentation and description*, vol. 4, 145-162. London: SOAS.
- Harris, Randy Allen. 1993. *The linguistics wars*. Oxford: Oxford University Press.
- Harrison, K. David. 2005. Ethnographically informed language documentation. In Peter K. Austin (ed.), *Language documentation and description*, vol. 3, 22-41. London: SOAS.
- Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie (eds.). 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
- Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11(1), 119-132.
- Haviland, John B. 2006. Documenting lexical knowledge. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics: Studies and Monographs 178), 129-162. Berlin: Mouton de Gruyter.
- Heine, Bernd & Derek Nurse (eds.). 2008. *A linguistic geography of Africa* (Cambridge Approaches to Language Contact). Cambridge: Cambridge University Press.
- Hellwig, Birgit & Friederike Lüpke. 2001. Caused positions. In Steve Levinson & Nick Enfield (eds.), *'Manual' for the field season 2001*, 126-128. Nijmegen: Max Planck Institute for Psycholinguistics, Language and Cognition Group.
- Hellwig, Birgit. 2006a. Semantics: Field work methods. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn. Oxford: Elsevier.
- Hellwig, Birgit. 2006b. Field semantics and grammar-writing: Stimuli-based techniques and the study of locative verbs. In Felix K. Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar-writing* (Trends in Linguistics: Studies and Monographs 167), 321-358. Berlin: Mouton de Gruyter.

- Hill, Jane H. 2006. The ethnography of language and language documentation. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics: Studies and Monographs 178), 113-128. Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36(1), 161-195.
- Himmelmann, Nikolaus P. 2006a. Language documentation: What it is and what it is good for? In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics: Studies and Monographs 178), 1-30. Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus P. 2006b. Prosody in language documentation. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics: Studies and Monographs 178), 163-181. Berlin: Mouton de Gruyter.
- Hymes, Dell. 1972. Models of the interaction of language and social life. In John J. Gumperz & Dell Hymes (eds.), *Directions in sociolinguistics: The ethnography of communication*, 35-71. New York: Holt, Rinehart and Winston.
- Kibrik, Aleksandr E. 2006. Collective field work: Advantages or disadvantages? *Studies in Language* 30(2), 259-279.
- Labov, William. 1975. *What is a linguistic fact?* (PdR Press Publications in Linguistic Theory 1). Lisse: Peter de Ridder Press.
- Ladefoged, Peter. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Blackwell.
- Levinson, Stephen & Sérgio Meira. 2003. 'Natural concepts' in the spatial topological domain – adpositional meanings in cross-linguistic perspective: An exercise in semantic typology. *Language* 79(3), 485-516.
- Levinson, Steve & Nick Enfield (eds.). 2001. *'Manual' for the field season 2001*. Nijmegen: Max Planck Institute for Psycholinguistics, Language and Cognition Group.
- Levinson, Stephen C. 2003. *Space in language and cognition: Explorations in cognitive diversity* (Language, Culture and Cognition 5). Cambridge: Cambridge University Press.
- Lüpke, Friederike. 2004. Language planning in West Africa – who writes the script? In Peter K. Austin (ed.), *Language documentation and description*, vol. 2, 90-107. London: SOAS.
- Lüpke, Friederike. 2005a. Small is beautiful: Contributions of small field-based corpora to different linguistic disciplines, illustrated by Jalonke. In Peter K. Austin (ed.), *Language documentation and description*, vol. 3, 75-105. London: SOAS.
- Lüpke, Friederike. 2005b. *A grammar of Jalonke argument structure* (Max Planck Series in Psycholinguistics 30). Ph.D. thesis, Radboud Universiteit Nijmegen.

- Maass, Anne & Aurore Russo. 2003. Directional bias in the mental representation of spatial events: Nature or culture? *Psychological Science* 14(4), 296-301.
- Majid, Asifa & Melissa Bowerman (eds.). 2007. Cutting and breaking events: A crosslinguistic perspective: Special issue of *Cognitive Linguistics* 18(2).
- Matras, Yaron, April McMahon & Nigel Vincent (eds.). 2006. *Linguistic areas: Convergence in historical and typological perspective*. Houndmills: Palgrave Macmillan.
- Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70(4), 369-415.
- Mithun, Marianne. 2001. Who shapes the record: The speaker and the linguist. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, 34-54. Cambridge: Cambridge University Press.
- Mosel, Ulrike. 2004. Dictionary making in endangered language communities. In Peter K. Austin (ed.), *Language documentation and description*, vol. 2, 39-54. London: SOAS.
- Mosel, Ulrike. 2006. Grammaticography: The art and craft of writing grammars. In Felix K. Ameka, Alan Dench & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar-writing* (Trends in Linguistics: Studies and Monographs 167), 41-68. Berlin: Mouton de Gruyter.
- Nathan, David & Peter K. Austin. 2004. Reconceiving metadata: Language documentation through thick and thin. In Peter K. Austin (ed.), *Language documentation and description*, vol. 2, 179-187. London: SOAS.
- Nathan, David. 2006. Thick interfaces: Mobilizing language documentation with multimedia. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics: Studies and Monographs 178), 363-379. Berlin: Mouton de Gruyter.
- Nathan, David. 2009. The soundness of documentation: Towards an epistemology for audio in documentary linguistics. In National Foreign Language Resource Center (ed.), *1st international conference on Language Documentation & Conservation (ICLDC)*. Honolulu: University of Hawai'i, National Foreign Language Resource Center. <http://hdl.handle.net/10125/5103> (8 February 2010).
- Newmeyer, Frederick J. 2007. Linguistic typology requires crosslinguistic formal categories. *Linguistic Typology* 11(1), 133-157.
- Ochs, Elinor. 1979. Planned and unplanned discourse. In Talmy Givón (ed.), *Discourse and syntax* (Syntax and Semantics 12), 51-80. New York: Academic Press.
- Özyürek, Asli, Sotaro Kita & Shanley Allen. 2001. *Tomato Man movies: Stimulus kit designed to elicit manner, path and causal constructions in motion events with regard to speech and gestures*. Nijmegen: Max Planck Institute for Psycholinguistics, Language and Cognition Group.

- Payne, Thomas E. & David J. Weber (eds.). 2007. *Perspectives on grammar writing* (Benjamins Current Topics 11). Amsterdam: John Benjamins.
- Raleigh Yow, Valerie. 2005. *Recording oral history: A guide for the humanities and social sciences*, 2nd edn. Oxford: AltaMira Press.
- Sapir, E. 1929. The status of linguistics as a science. *Language* 5(4), 207-214.
- Schegloff, Emanuel A. 2007. *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge: Cambridge University Press.
- Schiffrin, Deborah, Deborah Tannen & Heidi E. Hamilton (eds.). 2003. *The handbook of discourse analysis* (Blackwell Handbooks in Linguistics). Oxford: Blackwell.
- Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, IL: The University of Chicago Press.
- Schütze, Carson T. 2005. Thinking about what we are asking speakers to do. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 457-484. Berlin: Mouton de Gruyter.
- Scribner, Sylvia & Michael Cole. 1981. *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- Seifart, Frank. 2006. Orthography development. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics: Studies and Monographs 178), 275-299. Berlin: Mouton de Gruyter.
- Seifart, Frank. 2008. On the representativeness of language documentations. In Peter K. Austin (ed.), *Language documentation and description*, vol. 5, 60-76. London: SOAS.
- Sherzer, Joel. 2002. *Speech play and verbal art*. Austin, TX: University of Texas Press.
- Thieberger, Nicholas. 2004. Documentation in practice: Developing a linked media corpus of South Efate. In Peter K. Austin (ed.), *Language documentation and description*, vol. 2, 169-178. London: SOAS.
- Tomlin, Russell S. 1995. Focal attention, voice, and word order: An experimental, cross-linguistic study. In Pamela Downing & Michael Noonan (eds.), *Word order in discourse* (Typological Studies in Language 30), 517-554. Amsterdam: John Benjamins.
- van Staden, Miriam, Gunter Senft, Nick Enfield & Jürgen Bohnemeyer. 2001. Staged events. In Steve Levinson & Nick Enfield (eds.), *'Manual' for the field season 2001*, 115-125. Nijmegen: Max Planck Institute for Psycholinguistics, Language and Cognition Group.
- Van Valin, Robert D., Jr. & David P. Wilkins. 1996. The case for 'effector': Case roles, agents, and agency revisited. In Masayoshi Shibatani & Sandra A. Thompson (eds.), *Grammatical constructions: Their form and meaning*, 289-322. Oxford: Clarendon Press.
- Widlok, Thomas. 2005. Cross-disciplinarity in the documentation work of anthropologists and linguists. In Peter K. Austin (ed.), *Language documentation and description*, vol. 3, 12-21. London: SOAS.

- Wittenburg, P., U. Mosel & A. Dwyer. 2002. Methods of language documentation in the DOBES project. In Peter Austin, Helen Dry & Peter Wittenburg (eds.), *Proceedings of the international LREC workshop on resources and tools in field linguistics*, 36-42. Paris: European Language Resources Association.
- Woodbury, Anthony C. 2003. Defining documentary linguistics. In Peter K. Austin (ed.), *Language documentation and description*, vol. 1, 35-51. London: SOAS.
- Woodbury, Anthony C. 2007. On thick translation in linguistic documentation. In Peter K. Austin (ed.), *Language documentation and description*, vol. 4, 120-135. London: SOAS.
- Woodbury, Anthony C. & Nora C. England. 2004. Training speakers of indigenous languages of Latin America at a US university. In Peter K. Austin (ed.), *Language documentation and description*, vol. 2, 122-139. London: SOAS.
- Wooffitt, Robin. 2005. *Conversation analysis and discourse analysis: A comparative and critical introduction*. Thousand Oaks, CA: Sage.