

## Linguistic Data Types and the Interface between Language Documentation and Description

Nikolaus P. Himmelmann  
*Universität zu Köln*

This paper presents a new definition of documentary linguistics, based on a typology of linguistic data types. It clarifies the distinction between raw, primary, and structural data and argues that documentary linguistics is concerned with raw and primary data and their interrelationships, while descriptive linguistics is concerned with the relations between primary and structural data. The fact that primary data are of major concern in both fields reflects the fact that the two fields are very closely interlinked and difficult to separate in actual practice. The details of their interaction in actual practice, however, are still a matter for further discussion and investigation, as the second main part of the paper attempts to make clear.

**1. INTRODUCTION.**<sup>1</sup> Previous definitions of documentary linguistics have focused on its inter- and transdisciplinary nature. This has given rise to some misconceptions regarding its basis and role in linguistics, including the following:

- Documentary linguistics is all about technology and (digital) archiving.
- Documentary linguistics is just concerned with (mindlessly) collecting heaps of data without any concern for analysis and structure.
- Documentary linguistics is actually opposed to analysis.

If taken seriously, these views amount to the claim that documentary linguistics is not a linguistic enterprise, possibly not even a scientific one. In this paper, I argue that quite the opposite is true, i.e., that documentary linguistics is part of a much broader concern for putting linguistics on a proper empirical footing. In section 2, I present a systematics for linguistic data types according to processing stages and native speaker input, and then define the place of documentary linguistics with regard to this typology. As will be obvious from this discussion, documentary linguistics has the important task of making descriptive generalizations replicable and accountable, and in this sense it provides the empirical basis for many branches of linguistics.

---

<sup>1</sup>This paper is a revised version of a plenary presented at the 1<sup>st</sup> International Conference on Language Documentation & Conservation in Hawai'i in 2009. I am very grateful indeed to the organizers, in particular Nick Thieberger and Ken Rehg, for inviting me to this inspiring event and for making it such a wonderful experience. A first version of the first part of the paper was presented as a plenary at the annual meeting of the German Linguistic Society (DGfS) in Bielefeld in 2006. Dafydd Gibbon deserves special thanks for arranging this equally exciting occasion. I have profited considerably by the feedback from the audiences at both events, as well as at further presentations of parts of this paper in Bochum, Manchester, Thessaloniki, and, most recently, Cambridge University.

The current written version benefited significantly from the detailed and helpful comments offered by Sonja Gipper, Uta Reinöhl, Sonja Riesberg, and two anonymous reviewers for LD&C. Many thanks to all of you, and to Jessica Di Napoli for checking and improving grammar and style.

Still, the fact that language documentation and language description can be separated fairly clearly on methodological and epistemological grounds does not mean that they can be separated in actual practice. The complex nature of their interrelationship in this regard is still in need of further study and debate, as argued in section 3, which will be mainly concerned with the issues raised in Evans' (2008) review of Gippert et al. 2006.

## 2. LINGUISTIC DATA TYPES (according to processing stages and native speaker input).

While the last decade has witnessed an unprecedented interest in, and concern for, the empirical foundations of the discipline (cp. for example, Schütze 1996; Penke & Rosenbach 2004; Borsley 2005; Kepser & Reis 2005), to date only a few attempts have been made to provide a systematization of the basic data types used in structural linguistics, most importantly Iannàcaro (2000, 2001), Simone (2001:53 *passim*), and Lehmann (2004). Their systematizations are rooted in general epistemological and ontological distinctions, and we will briefly touch upon these distinctions at the end of section 2.1.

The starting point for the systematization proposed here, on the other hand, is the well-established and widely tested philological practice of the 18<sup>th</sup> and 19<sup>th</sup> centuries for dealing with historically attested language data, such as inscriptions or original manuscripts. In section 2.1, we will briefly review the major data types distinguished in this tradition and show that there are good methodological as well as ontological reasons for separating them in this way. In section 2.2 we will then show that the same basic distinctions may also be usefully applied to contemporary data.

**2.1 HISTORICAL LANGUAGE DATA.** In dealing with historically attested language data such as inscriptions or original manuscripts, philological practice distinguishes three basic types of data according to their processing stages (Table 1 provides a schematic overview).

	<b>Historical language data</b>
<b>RAW DATA</b>	<i>e.g., inscription, original manuscript(s)</i>
method(s) for deriving	(philological) criticism
<b>PRIMARY DATA</b>	<i>critical edition</i>
method(s) for deriving	historical-comparative + “interpretation”
<b>STRUCTURAL DATA</b> (secondary data)	<i>data on language and culture history</i> ( <i>e.g., Old High German o &gt; Middle High German ö</i> )

TABLE 1. Three processing stages for historical language data

The starting point for most philological enterprises is some kind of original written document. This original document, which we can call *raw data*, is often corrupt in one way or another. Thus, it may be incomplete, with a letter or a word missing in some places, or make use of abbreviations which are not immediately interpretable. Additionally, manuscripts often represent a text that was originally composed hundreds of years before the manuscript was written down (as is the case, for example, for all Ancient Greek philosophy and literature) and, in those instances in which a given original text (e.g.,

Aristotle's *Poetics*) has been transmitted in several manuscripts, there may be significant differences between the different versions.

In all these circumstances, the question arises as to what the original text/document actually looked like. Philology has developed a specific methodology, known as *philological criticism*, to address this issue. This is not the place to further expound this methodology,<sup>2</sup> the point of relevance here being simply that the application of this method to the original documents results in something known as a *critical edition* of the documents. Here, the editors present the text of an inscription or (set of related) manuscript(s) in what they consider to be the original version together with an apparatus which, among other things, contains explanations for all amendments that have been made to the original sources in determining the authoritative version published in the critical edition. Philological best practice requires that, as much as possible, the original sources be preserved and kept accessible so that later researchers who suspect that a particular amendment was based on an incorrect or incomplete understanding of the text can re-inspect the original source and propose a new reading and amendment of the segment in question.

The critical edition itself, however, is what usually serves as the basis for further research into the culture and language documented by the inscription or manuscript. Thus, for example, by comparing the forms of cognate words—as represented in critical editions—across centuries, it becomes possible to make observations such as that Old High German /o/ had become fronted /ö/ in the Middle High German period. The critical editions—not the raw data—provide the observational basis for generalizations about historical developments in a particular speech community on the understanding that the editors of the critical edition did their job well and provided the best possible reading of an original document. Someone interested in making observations regarding historical developments will thus (want to) look at the original document only in exceptional circumstances. In this sense, then, critical editions provide the **primary data** for observations and generalizations about historical developments in a given speech community.

Inasmuch as such observations or generalizations are considered to be true, they themselves become data (often called “facts”) for more general linguistic theorizing, providing (counter) evidence for theories of language change, the naturalness of phoneme inventories, etc. They could thus be called *secondary data* because they are two steps removed from the raw data (original sources). For reasons to be discussed shortly, this data type, however, will be called **structural data** here.<sup>3</sup>

In the preceding paragraphs, the distinction between raw, primary, and structural data is established in a “descriptive” way by simply recounting well-established practices in the

---

<sup>2</sup> The classic treatment is F.D.E. Schleiermacher 1977, which is a modern edition of a compilation of lecture notes first published in 1838. The lectures themselves were held between 1810 and 1830. A modern translation into English is *Hermeneutics and Criticism, And Other Writings*, translated and edited by Andrew Bowie, Cambridge University Press 1998. For a modern introduction to philological criticism, see West 1973.

<sup>3</sup> As an historical footnote, it may be of interest to note that in the early days of generative grammar a similar distinction was made between “data” and “facts” (Chomsky 1961:219). Later on, this distinction became blurred as part of a general move toward trivializing the empirical side of the discipline, as discussed, for example, in Labov 1975.

domains of philology and language and culture history. That these practices are not just contingent historical developments is indicated in Table 1 above through the inclusion of the major methods which are applied in between the three data types (or processing stages). That is, one systematic reason for distinguishing the three data types—and exactly these three data types—is a methodological one: each data type requires different methods. For example, the interpretation of raw historical data requires knowledge of scribal practices at the time of writing, some expertise in the physical properties of the media used for writing, hypotheses about authorship, etc. Such knowledge is of no, or at most rather peripheral, importance for detecting changes in linguistic structure. Establishing sound changes, for example, requires the compilation of cognate sets, hypotheses about possible and likely sound changes, etc.

Note that the claim that the three data types must be distinguished on methodological grounds does *not* mean that processing on each level can be done in total ignorance of the theories and methods relevant to the next level. This is not the case, as shown, for example, by the fact that whenever the precise nature of a postulated linguistic change is contested, it is not uncommon to find references to scribal practices (see Daunt 1939 and Stockwell & Barritt 1961 for a pertinent example from the history of English). The claim here merely asserts that each level requires specific methods relevant for handling the data type at hand, and that these methods have little role to play on other levels.

Apart from methodological reasons, there are also ontological and epistemological reasons to distinguish the three data types along the lines just indicated.<sup>4</sup> Raw and primary data are *particular*: i.e., they have a historical identity in the sense that they were produced at a specific point in time and space by a specific person (or group of persons), and this information is relevant in dealing with them.<sup>5</sup> Thus, for example, it matters who prepared a critical edition of a given set of inscriptions when, as is obvious from the fact that critical editions are attributed to specific editors and that for some influential works—e.g., the Bible—a number of different critical editions have been prepared over the centuries.

Secondary data are generalizations based on primary data (e.g., statements of sound changes in a given period). They are thus, by their very nature, *general* or, as we could also say, they lack a historical identity. In principle, it is not relevant who made a certain generalization (e.g., that Old High German /o/ became fronted /ö/ in Middle High German) or when this generalization was made. The only question that matters is whether this generalization is true in the sense of being corroborated/not falsified by the available primary data. If true, the generalization will remain true for as long as no conflicting primary data appears. Importantly, it will not change when another researcher works with the same set of primary data. That is, well-founded, robust generalizations based on

---

<sup>4</sup> The following terms and distinctions are taken from Lehmann 2004, but note that while Lehmann also distinguishes raw, primary, and secondary data, these distinctions are not fully commensurate with the ones made here. This is in part simply because Lehmann's systematization includes further distinctions not used here (e.g., raw vs. processed data).

<sup>5</sup> This does not mean that this information is always available. In fact, in the case of ancient manuscripts and inscriptions, it is often the case that it is not available, and it thus becomes part of the task of the philologist preparing a critical edition to provide hypotheses about the original place and time of authorship.

primary data should be replicable at any time by anyone sharing the same basic methods and assumptions used in making the generalization the first time around.

For this reason, it is common practice *not* to attribute secondary data to specific researchers, but instead to take them as “facts” having a standing of their own.<sup>6</sup> Only when they are controversial will generalizations based on primary data be attributed to their proponents (e.g., Rasmussen’s infix hypothesis for Proto-Indo-European (Rasmussen 1989)).<sup>7</sup> They will then also no longer be called “facts” but rather “hypotheses” or “claims.” To highlight the general and ahistorical nature of secondary data, they will be called *structural data* in the following.

The same type of replicability is not found in the relation between raw and primary data, because the derivation of primary data from raw data involves a considerable amount of interpretation and conjecture. There are no automatic, fully predictable procedures for this derivation, and hence primary data are non-unique: a team of editors working on the same set of raw data as another team will produce primary data (e.g., a critical edition) which will almost inevitably differ in some aspects from the primary data derived by the other team, even if the same methods and basic assumptions are applied (because interpretation and conjecture bring in subjective judgments). The differences are not necessarily substantial and may often be irrelevant for the generalizations to be based on the primary data (i.e., identical secondary data may be derivable from the two sets of primary data). But, as a matter of principle, there will always be such differences and hence there is no unique, fully determined set of primary data that can be derived from a given set of raw data. Instead, multiple derivations are possible (and, as mentioned above, have actually been carried out in cases such as manuscripts of the Bible).

This, then, constitutes a major ontological difference between raw and primary data: Raw data are unique. There cannot be two originals of the same inscription or manuscript. While it may be possible to reproduce originals in a number of ways, the results will always be copies, not originals. Primary data are non-unique in this sense, because it is always possible to derive another representation of a given set of raw data. But while critical editions of the same set of raw data are never identical, they are still alike in the sense of being representations of the same set of raw data.

Structural data are also non-unique, but in a somewhat different sense. On the one hand, they are replicable. On the other hand, because they are generalizations, the relation between them and the primary data they are based on is much less direct than the relation between raw and primary data. Hence, there are many different kinds of structural data that can be derived from primary data. Apart from sound change, which has been used as the main example of structural data throughout this section, historical primary data can form the basis for a broad range of observations and generalizations regarding not only all kinds of linguistic (grammatical, semantic) change, but also all kinds of cultural change. Further-

---

<sup>6</sup> The practice of historical linguistics to name some generalizations after the person who discovered them (e.g., Grimm’s Law) is only used in the case of particularly complex generalizations (e.g., a set of interrelated sound changes), thus crediting the originator’s outstanding insight. In addition, this practice is also due to the fact that it provides a simple way of referring to a complex generalization.

<sup>7</sup> Many thanks to Daniel Kölligan for providing this example.

more, historical primary data, of course, also have many synchronic uses in that they may serve as the basis for a grammatical description of the language used in the manuscripts or inscriptions, for analyzing ideas and attitudes prevalent in the society that produced the raw data, and so on. In order to distinguish the non-uniqueness of structural data from that of primary data, the attributes *replicable* and *variegated* will be used in reference to the former.

Table 2 summarizes the distinctions just introduced.

	Historical language data		Ontological properties
Raw data	e.g., <i>inscription</i>	=	particular, unique
Primary data	e.g., <i>critical edition</i>	=	particular, non-unique
Structural data	e.g., <i>sound law</i>	=	general/replicable (without historical identity), variegated

TABLE 2. Ontological properties of the three data types according to processing stage

**2.2 CONTEMPORARY DATA.** It is proposed here that the distinction between three basic levels of data processing (raw, primary, structural) can equally be applied to contemporary language data. The crucial difference between contemporary and historical data in the present systematization pertains to the fact that for contemporary data, direct speaker input is available. This is a crucial difference, because when native speakers can be involved in interpreting linguistic data, it is often possible to avoid at least some of the speculations involved in interpreting historical raw data.<sup>8</sup> Nevertheless, the availability of direct native speaker input does not mean that for contemporary data there is no distinction between raw and primary data, as we will see shortly.

It is useful to distinguish two types of contemporary language data according to the way native speakers are involved in their production. On the one hand, contemporary language data can be obtained by simply observing or recording communicative events (e.g., recording a conversation). This data type will be called *data based on observable linguistic behavior* in the following discussion. On the other hand, contemporary language data can be obtained by attempting to access native speakers' linguistic knowledge and skills more directly, for example, via elicitation or by carrying out a psycholinguistic experiment. This data type is characterized by the fact that native speakers are asked to engage in tasks which do not form part of their usual linguistic repertoires and often involve a reflective stance towards linguistic units or activities (as in providing acceptability judgments). This will be called *data based on metalinguistic skills*, and will be further discussed in section 2.2.2.

The distinction between data based on observable linguistic behavior and data based on metalinguistic skills is not a sharp one, in the sense that assigning every data specimen unambiguously to either type is not always a simple and straightforward task. Some

---

<sup>8</sup> Note that by this definition, historical language data do not have to be ancient. An unedited text handwritten by the last native speaker of a little known language last month becomes historical data in this sense the instant this speaker is no longer available for consultation.

data specimens may be hybrids, involving aspects of both data types. For example, if speakers are shown short video clips and then asked to describe the contents and classify the depicted action, the descriptions instantiate observable linguistic behavior, while the classification involves metalinguistic skills. However, the purpose of, and motivation for, the distinction is not to provide a neat classification grid for data specimens. This is likely not possible, because raw and primary data are usually messy regardless of whether they are of contemporary or historical origin. The purpose, rather, is to make it clear that data resulting from these two different kinds of activities require different kinds of processing methods, as will become obvious in the following two subsections.

<b>Raw data</b>	<i>recording (audio/video)</i>	<i>(non-recorded observation)</i>	<i>non-standard writing</i>
Methods for deriving	e.g., transcription, translation		“standardization”
<b>Primary data</b>	<i>transcript with translation</i>	<i>field notes</i>	<i>written document in standard orthography (edited non- standard text, newspaper, etc.)</i>
Methods for deriving	distributional and frequency analysis, tagging, cross-ling. comparison, “interpretation”		
<b>Structural data</b>	<i>descriptive statement, dictionary entry, interlinear glosses, frequency data, entry in typological database, treebank, implicational universal</i>		

TABLE 3. Different types of data based on observable linguistic behavior

**2.2.1 DATA BASED ON OBSERVABLE LINGUISTIC BEHAVIOR.** Table 3 lists the three major subtypes of data based on observable linguistic behavior. The best-known and prototypical data subtype of this sort are (audio or video) recordings of communicative events of any kind. While never providing a fully comprehensive record of a communicative event—even the most elaborate and ambitious set-up for video recording (with multiple cameras, etc.) can never fully match the total experience of a human observer present at the recorded event—audio and video recordings are still the best possible records of spoken linguistic behavior currently available. Importantly, they provide direct and persistent access to relevant aspects of a specific (original) communicative event which are otherwise of a rather ephemeral nature and hence difficult to capture. Before turning to the other two columns of Table 3, we will now first look more closely at the further processing of audio/video recordings, showing how the distinction between the three processing stages *raw*, *primary*, and *structural* applies to this data type .

Recordings are rarely used directly as the basis for further research, because the recorded event still remains ephemeral, and because a recording usually contains too much and too complex information. So, for linguistic purposes, it is standard practice to work with a transcript of the recording which, ideally, contains all and only the aspects of the recorded event relevant for a particular research project.

There are various styles of transcription currently in use in linguistics, including the type of transcripts used by conversation analysts and the type field linguists prepare when

working on a little-known language. To simplify the exposition, the latter will be used as the main example in the remainder of this article, but note that in principle similar observations and comments hold for other types of linguistic transcripts as well.

Transcription is by no means a trivial and straightforward exercise, as skilfully argued in a classic paper by Elinor Ochs (1979). There is no need here to repeat all the observations made in her paper. The important point for present purposes is that transcription aims to derive primary data (standardized symbolic representations) from raw data (observed linguistic behavior). This process involves segmentation on various levels, i.e., the identification of sound segments, words, and intonation units. While some evidence for these segmentation levels may be found in the recorded signal, a good transcription requires direct native speaker input and a hypothesis about the sound system and later also of the morphological structure of the variety being transcribed.

Native speaker input is particularly important with regard to the segmentation of words and phrases, because acoustic evidence for these units tends to be particularly weak. Obviously, a first indication of the meaning of lexical content items and the overall construction can also only be gained with the help of native speaker input. Consequently, the creation of a transcript of a recording requires the joint effort of someone who knows the language and someone who knows the principles of segmentation required for a useful symbolic representation of a speech event (this can be a single person in the case of a linguist working on her or his native language).

Segmentation and translation involve a certain amount of interpretation because neither is fully determined by the evidence available in the recording. As a consequence, two teams of researchers working on the same recording will not produce one hundred percent identical transcripts/translations (though, one would hope, that the two transcripts with translation would be reasonably similar and that the differences [for example, in representing clitic items] are irrelevant for many research purposes). In this regard, the relation between recordings and their transcripts and translations is similar to the relation between inscriptions or ancient manuscripts and a critical edition. For this reason, they are assigned to the same processing levels, i.e., raw and primary, respectively.

Like critical editions, transcripts with translations serve as the basis for structural generalizations and other types of structural data. As indicated in Table 3, structural data based on transcripts with translations can be of many different kinds, including:<sup>9</sup>

- descriptive statements, such as “the Waima’a particle *nini* marks possession with 3rd person possessors; *nini* always follows the possessum, but the possessor may precede or follow the possessum as in *ne wau nini* (3s pig POSS) = *wau ne nini* ‘her/his pig(s)’”;
- interlinear glosses which presuppose an analysis of grammatical and lexical structures and meaning;
- frequency statements, such as “in Waima’a postposed possessor constructions (e.g., *wau ne nini*) are less common than preposed ones (*ne wau nini*)”; and
- typological generalizations, such as “isolating languages tend to have serial verb constructions.”

---

<sup>9</sup> The Waima’a examples are based on the Waima’a documentation by Belo, et al. (2002–2006).

As indicated by the last example, structural data may differ significantly in their generality (a typological generalization has to be based on a cross-linguistic sample, and more often than not it is based on grammars rather than directly on textual data) and it may arguably be useful to attempt a further systematization of this large and heterogeneous category. But this is not a task for the present paper. In terms of data processing, structural data have the following important properties in common which justify their inclusion in a single (super-) category:

- They are all more or less directly based on primary data.
- The methods for deriving structural data from primary data allow for replicability—different people applying the same methods to the same data set should arrive at the same generalizations.
- Structural data are the kinds of data that feed directly into linguistic theorizing, i.e., empirical linguistic theories refer to structural data as their primary evidence.

That is, the structural data types listed in Table 3 all share the important ontological properties of being replicable and variegated (cp. Table 2), and may thus be put into one category with regard to processing stage.

Turning now to the other two columns of Table 3, we may note that recordings of communicative events are not the only kinds of data based on observable linguistic behavior. Aspects of communicative events can also be recorded by participant observers in the form of written field notes (Table 3, column 2). Such notes, however, differ radically from audio or video recordings in that they already employ symbolic representations. Hence, this data type already constitutes primary data for which the corresponding raw data are no longer accessible immediately after their occurrence. This, in turn, means that there is no way of verifying the accuracy of the notes. Of course, it will often be possible to verify that the recorded action, phrase, or gesture can actually be used in the kind of communicative event for which it has been recorded (and this is all that matters for a grammarian or ethnographer of communication). Still, what cannot be verified in any direct way is that the recorded action, phrase, or gesture actually occurred at the specific point in time and in the exact same manner as stated in the notes.

With regard to written manifestations of linguistic behavior, it should be noted that documents in standard orthography, such as contemporary books or newspapers, already constitute primary data as defined here. Most importantly, they already show multiple levels of segmentation (from paragraph to letter or sign) and adhere to known standards of representation. Thus, there is arguably no need for further editing before they can be used as primary data for structural analyses, as confirmed by current practices in corpus linguistics.<sup>10</sup>

The matter is different for documents written in a non-standard way (orthographically or in terms of punctuation) including, for example, handwritten notes, text messages (SMS),

---

<sup>10</sup> One might question whether the occasional emendation of errors in orthography or punctuation does not instantiate the processing step from raw to primary. The answer here depends on how automatic and free of subjective interpretation such emendations are.

online instant messaging and the like, which require specific interpretation in order to identify segments and determine intended meanings. More often than not, the interpretation required presupposes expert knowledge which either the writer or the social circle she or he belongs to may provide (as with, for example, text messages written in a currently in-vogue teen-speak or handwritten notes with lots of abbreviations known only to members of a particular group). Once this expert knowledge is no longer available, we are dealing with historical data in the sense that native speaker input may no longer be used and the primary data must now be derived from the raw data employing methods belonging to the realm of philological criticism.

<b>Raw data</b>	<i>reaction in an experiment (token), including acceptability ratings</i>	<i>elicitation of words, paradigms, taxonomies, etc.</i>	<i>(introspection by linguist)</i>
Methods for deriving	statistical analyses and tests	transcription, translation, statistics (?)	
<b>Primary data</b>	<i>statistically significant results of an experiment</i>	<i>field notes</i>	<i>invented example</i>
Methods for deriving	“interpretation,” distributional analysis, further statistical analysis, cross-linguistic comparison, etc.		
<b>Structural data</b>	<i>descriptive statements (including statements such as “construction X is ambiguous”), statements on differences in processing speed regarding two items or constructions, dictionary entry, interlinear glosses, frequency data, entry in typological database, treebank, implicational universal</i>		

TABLE 4. Different types of data based on metalinguistic skills

**2.2.2 DATA BASED ON METALINGUISTIC SKILLS.** The core feature of data based on metalinguistic skills is that speakers engage in linguistic activities which are not part of their usual linguistic repertoire. Table 4 lists the basic subtypes.

The defining feature of this data type—that speakers engage in linguistic activities which are not part of their usual linguistic repertoire—is perhaps most clearly illustrated by data coming from experiments. In line with the recent literature (e.g., Schütze 1996; Kepser & Reis 2005), this includes not only data from psycholinguistic experiments (e.g., reaction-time experiments on word recognition or sentence processing), but also all kinds of acceptability judgments in which speakers are asked to rate a linguistic unit presented to them in some way. Participating in a language-related experiment is obviously not part of any speaker’s everyday linguistic repertoire and involves the activation of metalinguistic skills, in the sense that linguistic knowledge and skills are deployed in a task that typically involves a reflective stance and the objectification of linguistic units.

This is, admittedly, a rather broad use of the term *metalinguistic*. Evans (2008:341) questions the appropriateness of this terminology, “since a more prototypical reading of the [term *metalinguistic*] focuses it on those aspects of language that overtly name and consciously theorize about language functions, meanings, and structures.” This “more prototypical,” but also fairly narrow, reading of the term is not intended here, but I do

not see any good alternatives and hence will continue to use it.<sup>11</sup> Importantly, observable linguistic behavior is, of course, also based on linguistic knowledge and skills, hence calling the second data type *data based on linguistic knowledge/skills* would be somewhat misleading and certainly not helpful in denoting the intended distinction.

Data from grammar and lexicon elicitation sessions, as they commonly occur in linguistic fieldwork, are certainly not experimental in the same sense as data from psycholinguistic experiments proper. Most importantly, perhaps, grammar and lexicon elicitation is usually not controlled for possible biases and usually does not involve statistical tests to assess the validity and relevance of the raw data (although, depending on the topic, such testing may actually be warranted; see section 3.2). It may also be argued that it is perhaps “less unnatural” than the proper experimentation in the sense that (some) speakers may be involved in activities of the type “how do you say X in your language?” or “is there a common name for all these types of plants?” in their everyday linguistic habitat (i.e., not as part of an interaction with a researcher). Still, elicitation focuses their attention on linguistic structures and practices in a way that requires introspection and objectification that is not part of everyday linguistic behavior. Furthermore, elicitation occurring in a research context is clearly more intensive than that occurring in everyday interactions, and it usually leads to the establishment of new routines on the part of the speakers, which is not unlike the routines of subjects repeatedly involved in linguistic experiments. In both types of data gathering, for example, problems may arise due to repetition or fatigue, as when speakers start reproducing the same type of structure when translating quite different input structures (elicitation) or develop an automatized routine in responding to stimuli (experiment).<sup>12</sup>

Elicitation and some types of experiments (such as acceptability ratings) build on introspection, which is one of the reasons why these data collection methods should usually be done with a number of different subjects so as to counter unwanted side effects of the nature of the task. One major unwanted influence on introspective judgments originates in theoretical biases. For this reason, linguists are generally not good subjects for data-generating methods involving introspection, as has been noted repeatedly over the last several decades (e.g., Labov 1975, Schütze 1996; both with many additional references). In-

---

<sup>11</sup> In previous work (e.g., Himmelmann 2006) data of this type was called “data based on metalinguistic knowledge.” Changing this to “data based on metalinguistic skills” is an attempt to make clear that a rather broad understanding of the term metalinguistic is intended here.

<sup>12</sup> One anonymous reviewer raises the issue of whether and how the practices covered by the broad definition of metalinguistic proposed here can be distinguished from the practices speakers engage in when interacting with small children acquiring a language, which surely should be considered part of (many) speakers’ usual linguistic repertoire. This is an intriguing issue in need of further consideration. While there are certainly similarities and overlaps between these two kinds of practices (as also hinted at in the paragraph above), there are also clear differences with regard to intensity, reflective stance, objectification of linguistic units, and participant structure (adult–child, typically in kin relation, vs. adults who interact primarily in order to document linguistic structures and practices). It is highly likely that at least in some cultures and societies, metalinguistic skills displayed in linguistic elicitation and experimentation build on everyday practices in adult-child interaction. But I still believe that the differences are significant enough not to include them in the same category.

vented examples based solely on the intuition of the linguist in order to support a certain theoretical point are now widely discarded as unacceptable linguistic data types.<sup>13</sup> In terms of the typology developed here, they are problematic as data because the raw data—and the methods for deriving primary data from them—are not accessible and hence lack accountability.

This does not necessarily mean that invented examples must never be used in linguistic argumentation. They may still be legitimate shortcuts in unproblematic “clear cases,” as they are sometimes called. But then, what are “clear cases?” We could consider them to be typically high-frequency constructions for which real examples can easily be extracted from spontaneous corpora (the English article-noun construction [*the child*] being a classic example). See also the discussion of elicitation in section 3.2.

**2.3 DATA TYPES AND THE DISTINCTION BETWEEN DOCUMENTARY AND DESCRIPTIVE LINGUISTICS.** Table 5 summarizes the data types discussed in this section. There are two major parameters: 1) the processing stage, with the three basic stages *raw*, *primary* and *secondary*; and 2) the way in which native speaker input is accessible. With regard to the latter, the major difference is between historical data for which native speaker input is no longer available and contemporary data, with which it is still possible to have native speakers generate new data or explain or evaluate already-collected data. The distinction between the two major types of contemporary data, i.e., those based on observable linguistic behavior and those involving the deployment of metalinguistic skills, is a gradual one, with many data specimens involving aspects of both kinds.

---

<sup>13</sup> In fact, and even more generally, the widely-made distinction between “grammaticality judgments” and “acceptability judgments” has been called into question in terms of the levels of raw and primary data by Schütze (1996:26), who convincingly argues: “It does not make any sense to speak of grammaticality judgments given Chomsky’s definitions, because people are incapable of judging grammaticality—it is not accessible to their intuitions ... . Linguists might construct arguments about the grammaticality of a sentence, but all that a linguistically naïve subject can do is judge its acceptability.”

	Direct input from native speaker		Indirect input from native speaker
	Data based on <b>observable linguistic behavior</b>	Data based on <b>metalinguistic skills</b>	<b>Historical data</b>
<b>Raw data</b>	e.g., <i>recording of a discussion</i>	e.g., <i>acceptability rating (token)</i>	e.g., <i>inscription, original manuscript</i>
Method(s) for deriving	e.g., transcription and translation	e.g., statistical analyses	(philological) criticism
<b>Primary data</b>	e.g., <i>transcript (with) translation</i>	e.g., <i>acceptability rating (statistics)</i>	<i>critical edition</i>
Method(s) for deriving	e.g., distributional and frequency analysis, tagging, cross-linguistic comparison, interpretation	statistical testing and modelling, interpretation	historical-comparative + “interpretation”
<b>Structural data</b> (secondary data, a.k.a “facts”)	<i>descriptive statements, dictionary entry, interlinear glosses, frequency data, typological databases, treebank, implicational universal</i>		<i>data on language and culture history (e.g., OHG o &gt; MHG ö)</i>

TABLE 5. Basic linguistic data types according to native speaker input (columns) and processing stage (rows)

This typology provides for another way to delimit documentary and descriptive linguistics, complementing and refining the earlier proposal in Himmelmann 2004: 39–47.<sup>14</sup> Documentary linguistics (dotted border in Table 5) is primarily concerned with raw and primary data and their interrelationships, including issues such as the best ways for capturing and archiving raw data, transcription, native speaker translation, etc. Descriptive linguistics (bold border), on the other hand, deals with primary and structural data and their interrelationships, i.e., primarily with the question of how valid descriptive generalizations can be derived from a set of primary data. Primary data (gray shading) thus have a dual role, functioning as a kind of hinge between raw and structural data. They are the result of preparing raw data for further analysis (documentation), and they serve as input for analytical generalizations (description). Only when primary data and the raw data on which they are based are made available will it be possible to check and replicate descriptive generalizations. In this view, documentation has the central task of making description accountable and replicable, and is thus of fundamental importance for making linguistics an empirical science.

<sup>14</sup> A condensed version of the earlier proposal can be found in Himmelmann 1998:161–164.

In the preceding paragraph, the term *documentary linguistics* is given a rather broad definition. In current usage, one may also find a somewhat narrower interpretation which only refers to raw and primary data based on observable linguistic behavior (column 1 in Table 5), or even narrower than that, to such data gathered in fieldwork in small, non-western communities. Work explicitly concerned with the collection of raw and primary data based on metalinguistic skills is currently often presented under the label of *experimental linguistics*. As the two data types overlap and are not clearly separable in many instances, it is doubtful that such a distinction would really be helpful.

**3. THE INTERFACE BETWEEN DESCRIPTION AND DOCUMENTATION IN ACTUAL PRACTICE.** As emphasized repeatedly in previous work (Himmelmann 1998:162f, 2006:28), the conceptual separation of language documentation from language description does not mean that these two scientific projects and activities are separable in actual practice, i.e., that one can be done without the other. Language documentation necessarily involves language description inasmuch as description provides basic input for major documentary activities such as transcription (practical orthography, word and phrase segmentation) and for deciding on what to document (the accounting function of analysis, which is further discussed in section 3.3 below). And, likewise, descriptive analysis clearly needs a corpus of primary data about which to make generalizations. Furthermore, as argued in the previous section, raw and primary data have to be properly curated and made accessible so that descriptive generalizations can be tested and replicated.

However, while the principle inseparability of documentation and description in actual practice is not contested seriously anywhere, there appear to be a few controversial or misconceived points as to precisely how the interrelationship between documentation and description is to be conceptualized and, perhaps more importantly, carried out in actual practice. In recent discussions of the separability issue, most notably Evans 2008: 346–348, Chelliah & de Reuse 2011:11–17, and Woodbury 2011:177f, three closely related issues tend to reappear, but again, they are best kept separate for the sake of conceptual clarity. These issues all revolve around the problem of setting priorities, in particular setting priorities in practice. They are:

- 1) The question of how to resolve the tension between ideas of what an ideal documentation could and should look like and the grim realities of constrained resources (see section 3.1).
- 2) The role of grammar-targeted elicitation in language documentation and description, which is often seen to be in opposition to work done with, and on, natural discourse data (see section 3.2).
- 3) The role of fully worked-out descriptive grammars in language documentation (see section 3.3).

**3.1 “PIE IN THE SKY” AND THE DIVERSITY OF DOCUMENTATION PROJECTS.** It would seem to be uncontroversial that, from a purely theoretical point of view, abstracting away from the exigencies of the real world of constrained resources, every language should be documented and described to the fullest degree conceivable. This would comprise a massive corpus of primary data including not only recorded communicative

events, but also elicited and experimental data of different kinds, and for all topics of interest. It would also comprise not only a single descriptive grammar and a dictionary, but a number of descriptive analyses and dictionaries in different formats and with differing emphases: a pan-dialectal grammar highlighting intragroup variation vs. one focusing on a single variety, a semasiologically organized grammar and an onomasiologically organized one, topical dictionaries, etc.

But this ideal is “pie in the sky,” as Chelliah & de Reuse (2011:13) aptly note in their discussion of the separability issue. Realistically speaking, when working on un(der)-documented languages, resources are always limited and hence priorities have to be defined whenever one goes beyond the basic steps.<sup>15</sup> It is at this point that conflicts of interest may arise between the different tasks and goals that language documentation and description in principle demand. The potential for such conflicts naturally increases to the extent that the documentation component also takes on board interests and demands from non-linguist user groups, such as other academic disciplines and the speech community itself.

I do not see a way to a principled resolution of such conflicts of interest on theoretical grounds, as this would involve weighing the interests of one user group over those of another, and answering questions such as: Would it be possible, and does it make sense, to argue that a descriptive grammar is—on principled grounds—more important than an ethnography or a comprehensive dictionary, or than some other materials demanded by the speech community?

Inasmuch as one agrees that the answer to this type of question has to be *no*, it is clear that there can only be a pragmatic resolution of conflicts of interest arising in language documentation and description, along the lines of the following principle.

**Pragmatic resolution of conflicts of interest in language documentation:**

Do what is pragmatically feasible in terms of the wishes and needs of the speech community and in terms of your own specific skills, needs, and interests.

That is, it does not make sense to demand that a linguist write an ethnography, or that someone interested in lexical semantics struggle with eliciting data on control constructions. Typologically-minded linguists will want to write a descriptive grammar or series of papers on structural phenomena of interest from a cross-linguistic perspective. It would be naïve and wrong to assume that researchers do not have their own interests and agendas when engaging in documentary activities; and while these need to be checked and balanced against the interests of other stakeholders (the community, the funding agency, the discipline, the wider public, etc.), it is legitimate and simply rational that everyone engage as much as possible in whatever they like doing and thus, as a rule, tend to do best. As Dobrin et al. have noted:

---

<sup>15</sup> That is, the first set of recordings and elicitation sessions (sociolinguistic and grammatical) needed for a first basic analysis (sketch grammar, practical orthography based on phonemic analysis) and for getting an idea of how the speech community is organized, what types of communicative events would probably need to be documented, and which grammatical, lexical or sociolinguistic topics need further study.

Each research situation is unique, and documentary work derives its quality from its appropriateness to the particularities of that situation.... Rather than approaching endangered languages with preformulated standards deriving from their own culture, documentary linguists must strive to be singularly responsive—both to what is distinctive about each language as an object of research, and to the particular culture, needs, and dispositions of the speaker communities with whom their work brings them into contact.... (Dobrin, et al. 2009:47)

Clearly, the skills and interests of the members of the documentation team add one more component to the uniqueness of each documentation project.

**3.2 ON DIVIDING RESOURCES BETWEEN WORK ON NATURALISTIC DATA AND GRAMMAR-TARGETED ELICITATION.** A recurrent misunderstanding with regard to documentary linguistics pertains to the idea that documentation should focus first and foremost on discourse data (i.e., more or less natural data documenting linguistic behavior) and that elicitation should be avoided altogether. This is wrong, as the above discussion in section 2 should have made clear. Data based on metalinguistic skills are of central importance to understanding the workings of a language, and collecting these data necessarily involves elicitation techniques of various kinds (including psycholinguistic experiments). Hence, elicitation (broadly understood) is necessarily a part of any documentation project that even remotely aspires to be comprehensive (unattainable as full comprehensiveness may be).

A related, but somewhat different, issue is the question of to what extent elicitation is necessary in order to make available all the data needed for a full grammatical analysis of the variety being documented. The answer here depends on what kinds of data are deemed to be insufficiently represented, even in large corpora covering a broad range of differing communicative events. One obvious data type missing from corpora is negative evidence of the kind provided by judgments of (un)acceptability. That is, it is widely assumed that the explicit rejection of an invented example by native speakers is stronger evidence for the fact that examples of this type are not possible in the language than the evidence provided by the fact that examples of this type are not to be found in a corpus, as this could always be due to chance.

But there are also other types of data which tend to be absent from corpora, usually the kind of data needed for complex grammatical analyses. Evans gives the following example:

In the realm of phonology, Hyman (2007), shows just how outrageously unnatural are the N+N+N combinations one needs to elicit in order to work through all the possible tone combinations needed to plumb the depths of tone sandhi in Kuki-Chin languages. To check out all the combinations of floating tones that are needed to test particular hypotheses, it is necessary to construct sequences like ‘chief’s beetle’s kidney basket’ or ‘monkey’s enemy’s snake’s ear’. (Evans 2008:347)

While I agree that there are some topics of grammatical analysis where elicitation may provide the evidence that cannot be found in a comprehensive corpus, it seems to me that discussions of this issue tend to miss the following crucial point: structures not occurring with some frequency in everyday, natural speech are generally also not easy to elicit. Consequently, elicitation is often a productive strategy in instances in which it is not possible to compile a truly comprehensive corpus of natural speech (which is, of course, the rule rather than the exception). But note that elicitation may not be very productive even in cases of high frequency items, discourse particles being a well-known case in point.

More importantly, when elicitation targets expressions that rarely, if ever, occur in natural speech, it requires considerably higher effort and methodological rigor than is typically applied in order to produce robust results. If it is indeed necessary to elicit “outrageously unnatural” examples such as *chief’s beetle’s kidney basket* in order to check all the possible tone combinations in Kuki-Chin languages, this goal cannot be achieved by simply asking one speaker one time about the acceptability of the example. Given the unnaturalness of the examples, one needs to show that speakers across the community behave consistently with regard to them. This, in turn, implies a rigid sampling and testing procedure<sup>16</sup> which, among other things, should include testing the same examples with the same speakers several times. Typically, at least in my experience and as now documented in the literature on experimental linguistics referred to above, speakers do not react in a uniform and easily interpretable way when confronted with such examples. Consequently, their reactions need considerable interpretation, including statistical analyses, in order to be useful for further analysis.

In short, I do not believe that there are simple guidelines with regard to the issue of how to divide limited time and resources between corpus collection and annotation and more specialized elicitation, especially elicitation targeted at grammatical topics.<sup>17</sup> To a large degree, this will depend on the skills and interests of the parties involved: some speakers, and some linguists, are better at corpus work, while others are more productive in specialized elicitation.

**3.3 THE ROLE OF FULLY WORKED-OUT DESCRIPTIVE GRAMMARS IN LANGUAGE DOCUMENTATION.** Given the close interrelationship and the many interdependencies between documentation and description, the question arises as to whether there is a real difference between projects flagged by one or the other label. In my opinion, there is indeed

---

<sup>16</sup> Somewhat surprisingly, while Hyman actually concludes the paper with the assertion, “The experimental nature of elicitation should therefore not be underestimated,” he does not address the issue of what this means for gathering and validating data.

<sup>17</sup> Phrasing the alternative this way is intended to emphasize the fact that work on a useful corpus does not, of course, simply consist of the collection of recordings. Instead, it also comprises work on annotating the recordings, which in turn will include various kinds of elicitation that are contextualized by the recording one is working on. In this regard, I fully concur with Evans’ (2008:347) statement that “to be really useful a corpus must contain discussions of the various ways that each sentence in it can be interpreted in different contexts—a sort of semantically annotated meta-corpus. Again, this can only be produced by embroidering unstructured text with elicited probings—what if you had said X instead? what would it have meant? and so forth.”

a difference, though it is primarily one of emphasis and perspective. As argued in previous work (Himmelmann 1998, 2004), in documentation, description has an ancillary function, and vice versa.

If one takes seriously the position that description has a subsidiary function in documentation, it requires us to scrutinize established descriptive formats such as reference grammars, asking whether they serve this function in the best possible way. This question has not been explored to date to any appreciable degree. The present section does not aim to make a major contribution in this regard either. Rather, its primary purpose is to provide further support for the view that this is indeed a topic worthy of further scrutiny.

What is clear and relatively widely agreed upon, is that the ancillary function of description in documentation has two major aspects. First, description serves an *accessibility function*, as analysis is needed to make the data compiled in a documentation accessible in a useful way (thereby avoiding dilemmas commonly presented by unannotated raw data). On the most basic level, this comprises the analyses necessary to make useful transcriptions and free translations, and to provide a glossary of all items found in a given corpus. This may be extended to include interlinear glosses and their explanations. Whether or not the accessibility function also demands a full-fledged descriptive grammar is an issue that requires further discussion and exploration. Clearly, it would have to be conceived of as a tool for accessing and understanding the materials in the corpus rather than being primarily oriented towards the discourse of typologists, as most current reference grammars are. Heath's (1984) Nunggubuyu grammar may provide a starting point for the further development of a grammar format serving this function.

Of equal importance is the *accounting function* of analysis (cp. Rhodes, et al. 2006, quoted in Evans 2008:346f<sup>18</sup>). The major function of producing comprehensive analyses in this view is to ensure the comprehensiveness of the documentation. That is, analysing the data is one, if not the major, means for diagnosing "holes" in the data. Evans argues that this function is best served by the well-established format of descriptive reference grammars:

The most important challenge faced by one writing a reference grammar is to construct a description whose thousands of sub-analyses capture the overall Bauplan of the language and at the same time succeed in being mutually consistent. To see this as mere formulation and organization is to grossly underestimate the nature of the analytic challenge. Just having a series of analytic sections that hang off the documentary material runs the triple risk of incompletely pursuing the specific analytic questions, failing to pick up on the interactions between different subanalyses, and representing the analytic claims as a miscellaneous catalogue rather than an organized whole. (Evans 2008:348)

---

<sup>18</sup> This paper is cited as follows in Evans 2008; it is not available to me.

Rhodes, Richard A., Lenore A. Grenoble, Anna Berge, and Paula Radetzky. 2006. *Adequacy of documentation: A preliminary report to the CELP* [Committee on Endangered Languages and their Preservation, Linguistic Society of America].

Bringing together different analyses and trying to present them as a coherent whole is certainly one way of discovering inconsistencies and missing parts, the removal of which may require gathering additional data. However, this venue for fully discharging the accounting function of analysis is not without its problems. To begin with, it seems to overstate the overall coherence of linguistic systems. It is doubtful that there are indeed anything like “great underlying groundplans” (Sapir 1921:144) comprising all aspects of a given system, from segmental phonology to complex sentence structure, including numeral systems, compounding, interjections, comparatives, etc. (cp. Comrie 1989:40–42 *passim* for a rather pessimistic assessment of the prospects for holistic typologies). Hence, consistency checks provided by trying to form a coherent whole out of different partial analyses hold primarily for the level of subsystems rather than for the language as a whole.

More important is the problem of practical feasibility: how realistic is it to expect to be able to write a reference grammar of the quality and comprehensiveness envisioned in the above quote within the constraints of a documentation project? Current practice shows that projects 3–5 years in duration produce either a reasonably comprehensive corpus with some analytical papers or a full reference grammar, but not both. This raises the question of whether there are other, possibly more efficient methods than reference grammars for discharging the accountability function of descriptive analysis within the context of language documentations proper. (See Thieberger 2009 for some very preliminary ideas on this important topic in need of further investigation within documentary linguistics.)

Finally, it should be emphasized that while the discussion of the role and extent of description in documentation is an important topic in the theory of language documentation, part of the dynamics of the field derives from the fact that this is not an issue which needs to be settled before actual useful work can be done. That is, even if one agrees with the position that comprehensive reference grammars are the best option for accomplishing the accessibility and accountability functions of descriptive analysis within a documentation project (from a linguistic point of view), this does not mean that writing a reference grammar should become a core requirement for documentation projects. It still makes sense to pursue documentation even in contexts where no one is willing or able to write a reference grammar, for the simple reason that imperfect documentation is still better than no documentation. In fact, very useful contributions to language documentation can be made without engaging in grammar writing, as text collections and collections of specialist terminologies provide a wealth of interesting data even if they turn out to be incomplete with regard to the goal of writing a reference grammar.

However, this does not mean that as a rule documentation projects should not strive to produce some published product. As part of his plea for a central role of fully worked-out reference grammars in language documentation, Evans also observes:

[T]he degree of native-speaker involvement and critical engagement increases dramatically at the point where a published product is prepared (say a bilingual edition of a story for school uses, or a first dictionary of a language). Something about the definitive appearance of these products brings out a higher level of scrutiny and a leap to new levels of accuracy in transcription and translation. Both times that I have been involved in producing dictionaries of Australian Aboriginal languages, [FN omitted, NPH] there was a sudden upsurge in interest and in the

supplying of new or extended lexical entries at the point where speakers of the language held in their hands a properly-produced book in their language. (Evans 2008:348)

This is a point which I find myself in full agreement with and which, to my mind, has not been sufficiently emphasized in the writings on language documentation. It should become a feature of a typical (there are always exceptions!; see section 3.1) documentation project to produce some type of publication other than the archival materials usually forming the core concern of such a project.

#### REFERENCES

- Belo, Maurício C.A., John Bowden, John Hajek, Nikolaus P. Himmelmann & Alexandre V. Tilman. 2002–2006. *DoBeS Waima'a documentation*. DoBeS Archive MPI Nijmegen, <http://www.mpi.nl/DOBES>.
- Borsley, Robert D. (ed.). 2005. *Data in theoretical linguistics*, Special issue of *Lingua* 115. 1475–1665.
- Chelliah, Shobhana L. & Willem J. de Reuse. 2011. *Handbook of Descriptive Linguistic Fieldwork*. London: Springer.
- Chomsky, Noam. 1961. Some methodological remarks on generative grammar. *Word* 17. 219–239.
- Comrie, Bernard. 1989. *Language universals and linguistic typology*, 2<sup>nd</sup> edn. Chicago: The University of Chicago Press.
- Daunt, Marjorie. 1939. Old English sound changes reconsidered in relation to scribal tradition and practice. *Transactions of the Philological Society* 38. 108–37.
- Dobrin, Lise M., Peter K. Austin & David Nathan. 2009. Dying to be counted: The commodification of endangered languages in documentary linguistics. In Peter K. Austin (ed.), *Language Documentation and Description*, vol. 3, 37–52. London: SOAS.
- Evans, Nicholas. 2008. Review of Gippert et al. 2006. *Language Documentation & Conservation* 2. 340–350.
- Gippert, Jost, Nikolaus P. Himmelmann & Ulrike Mosel (eds.). 2006. *Essentials of language documentation*. Berlin: Mouton de Gruyter.
- Heath, Jeffrey. 1984. *Functional grammar of Nunggubuyu*. Canberra: AIAS.
- Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36. 161–195.
- Himmelmann, Nikolaus P. 2004. Documentary and descriptive linguistics. In Osamu Sakiyama & Fubito Endo (eds.), *Lectures on endangered languages 5: from the Tokyo and Kyoto Conferences 2002*, 37–83 [= extended version of Himmelmann 1998].
- Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In J. Gippert, N. P. Himmelmann & U. Mosel (eds.), 1–30.
- Hyman, Larry M. 2007. Elicitation as experimental phonology: Thlantang Lai tonology. In Maria-Josep Solé, Pam Beddor & Manjari Ohala (eds.), *Experimental approaches to phonology in honor of John J. Ohala*, 7–24. Oxford: Oxford University Press.

- Iannàcaro, Gabriele. 2000. Per una semantica più puntuale del concetto di ‘dato linguistico’: Un tentativo di sistematizzazione epistemologica. *Quaderni di Semantica* 21. 51–79.
- Iannàcaro, Gabriele. 2001. Alla ricerca del dato. In Federico Albano Leoni, Rosana Sornicola, Eleonora Stenta Krosbakken & Carolina Stromboli (eds.), *Dato empirici e teorie linguistiche*. Atti del XXXIII Congresso Internazionale di Studi della Società Linguistica Italiana, 23–35. Roma: Bulzoni (SLI 43).
- Kepser, Stephan & Marga Reis (eds.). 2005. *Evidence in linguistics*. Berlin: Mouton de Gruyter.
- Labov, William. 1975. *What is a Linguistic Fact?* Lisse: de Ridder.
- Lehmann, Christian. 2004. Data in linguistics. *The Linguistic Review* 21. 175–210.
- Ochs, Elinor. 1979. Transcription as theory. In Elinor Ochs & Bambi B. Schieffelin (eds.), *Developmental pragmatics*, 43–72. New York: Academic Press.
- Penke, Martina & Anette Rosenbach. 2004. What counts as evidence in linguistics? *Studies in Language* 28(3). 480–526.
- Rasmussen, Jens Elmegård. 1989. *Studien zur Morphophonemik der indogermanischen Grundsprache*. Innsbrucker Beiträge zur Sprachwissenschaft 55. Innsbruck: Institut für Sprachwissenschaft.
- Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace & Company.
- Schleiermacher, Friedrich D.E. 1977. *Hermeneutik und Kritik*. Frankfurt am Main: Suhrkamp.
- Schütze, Carson T. 1996. *The empirical base of linguistics*. Chicago: The University of Chicago Press.
- Simone, Raffaele. 2001. Sull’utilità e il danno della storia della linguistica. In Giovanna Massariello Merzagora (ed.), *Storia del pensiero linguistico: Linearità, fratture e circolarità*. Atti del Convegno della Società Italiana di Glottologia, Verona, 11–13 novembre 1999, 45–67. Roma: il Calamo.
- Stockwell, Robert P. & C. Westbrook Barritt. 1961. Scribal practice: Some assumptions. *Language* 37. 372–389.
- Thieberger, Nicholas. 2009. Steps toward a grammar embedded in data. In Epps, Patricia & Alexandre Arkhipov (eds.), *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, 389–408. Berlin & New York: Mouton de Gruyter.
- West, Martin L. 1973. *Textual criticism and editorial technique*. Stuttgart: Teubner.
- Woodbury, Anthony C. 2011. Language documentation. In Peter K. Austin & Julia Salbank (eds.), *The Cambridge Handbook of Endangered Languages*, 159–186. Cambridge: Cambridge University Press.

Nikolaus P. Himmelmann  
sprachwissenschaft@uni-koeln.de