Nikolaus P. Himmelmann

What about typology is useful for language documentation?

DOI 10.1515/lingty-2016-0020
Received July 13, 2016; revised September 21, 2016

1 Language documentation and language description

Before addressing the question in the title, it will be useful to briefly clarify what is meant by "language documentation", as this term is currently used in two senses. In its broader use, it encompasses both the collection and initial processing of linguistic data as well as their analysis in the formats of a (descriptive) grammar and a dictionary. In its narrower sense, it focuses on data collection and processing, processing being concerned with transcription and translation as well as with making the data available for researchers and other interested parties, including the speech community.

In actual practice, the collection and initial processing of linguistic data and their descriptive analysis are inseparable activities. The proper and useful representation of linguistic data requires descriptive analysis (transcription, translation). And, vice versa, descriptive analysis needs data, the quality of the analysis depending to a substantial degree on the quality of the data. However, it is useful to keep these two activities separate CONCEPTUALLY, as they differ in their methods and their primary outcomes: a corpus of annotated primary data on the one hand, and a grammar-cum-dictionary on the other. More technically (Himmelmann 2012), language documentation is concerned with the collection of raw data, i.e., (audiovisual) recordings, and their representation as primary data, i.e., as transcripts with translations. Fieldnotes are a type of primary data delinked from raw data (= native speaker replies in an elicitation session). Language description, on the other hand, is concerned with deriving structural data from primary data, i.e., descriptive generalizations across a corpus of primary data.

The conceptual independence of documentation and description has practical ramifications. Thus, the emergence in the last two decades of language

Nikolaus P. Himmelmann, Universität zu Köln, Allgemeine Sprachwissenschaft, Institut für Linguistik, 50923 Köln, Germany, E-mail: sprachwissenschaft@uni-koeln.de

documentation proper as a field of linguistic inquiry in its own right has led to an increased concern for aspects of linguistic fieldwork that had earlier tended to be neglected. To give but two examples. First, there is now much more reflection on the question of how to involve native speakers in the creation of a lasting record of their language and how to make this useful for the speech community. Second, the issue of properly archiving raw and primary data in a way that allows for further use in the long-term future has become a major industry in the field.

A core issue of the documentarist agenda that is still insufficiently addressed pertains to (discourse) transcription and translation: (i) how to segment and represent specimens of spoken language above the segmental level; (ii) how to represent the meaning of elicited and recorded materials in such a way that is useful for further analysis. We will return to the latter issue below.

In the remainder, "language documentation" will be used in its narrow sense, with "language description" referring to grammar and dictionary making. As the usefulness of typology for descriptive linguistics is the primary topic of other contributions in this issue of *LT*, an attempt is made here to discuss points that are specific to language documentation (narrowly conceived). But note that most points mentioned for language description in this regard also apply to language documentation. The two publications that have previously explored the relationship between language documentation and linguistic typology (Bond 2010; Epps 2010) make use of the broader notion, i.e., they do not distinguish between the two fields/activities.

2 The usefulness of typology for language documentation (and description)

Perhaps the most basic point in this regard is the fact that typology provides a *raison d'être* for language documentation. One of the purposes of language documentations is to serve as a primary data pool for typological inquiry. However, it is probably fair to say that typology still has a long way to go in order to make full use of this data source, as currently the main data pool still consists of structural (descriptive) data, often further abstracted to fit into a database scheme. The challenge consists in finding ways to extract relevant data from the primary (and perhaps even the raw) data which make up the core of a language documentation. Recent typological work on parallel texts (Cysouw & Wälchli 2007; see also http://paralleltext.info/) and in particular work by the Nijmegen Language and Cognition group using conversational and stimulus-

based data (e.g., Enfield et al. 2010; Dingemanse et al. 2013) provide examples of how such work could look.

This brings us to a second major point where typology provides important input for language documentation. Language documentation is in need of tools for generating (useful) data. Here, questionnaires and, in particular, the more recently developed picture and video stimulus toolkits have been found to be very useful (at least in my experience). Importantly, it may be the case (and in fact is perhaps more often the case than not) that these elicitation tools do NOT generate the kind of data they were designed for. To give a classic example, the well-known *Frog story* does not necessarily generate narratives, but may lead to unconnected picture descriptions. Or, video clips produced to generate reciprocal constructions may generate responses reflecting on the clothes of the participants or the natural setting of a scene. From a purely documentarist point of view, however, such reactions are not "failures" as they usually still produce interesting data (such as picture descriptions, constructions profiling landmarks, etc.).

It should be obvious that typology is particularly useful for language DESCRIPTION by making researchers aware of the diversity of possible linguistic structures and by providing examples of phenomena that may be of use in analyzing a given language (as has repeatedly been noted by Bond (2010), Epps (2010), and many others). While perhaps less directly so, the same also applies to language DOCUMENTATION inasmuch as it involves analysis. By necessity, language documenters have to engage with phonological analysis (to develop a useful and well-grounded practical orthography), but of course all elicitation, as well as the annotation of textual data, is greatly enhanced by being familiar with grammatical structures and discursive strategies found in other languages of the world.

Recently, it has become fashionable to emphasize the difference between language description and typological comparison, with authors such as Lazard (2002), Croft (2001), and Haspelmath (2010) actually conceiving of language description as an enterprise completely separate from typological and, it would seem, all other nondescriptive linguistic inquiry. For practitioners of language description, I would think, this view is ill-conceived and does not reflect current practice, which values the typologically informed model of a descriptive grammar most highly. Perhaps most importantly, descriptivists and typologists work with the same set of analytical techniques and concepts, including complementary distribution, articulation place features, semantic scope effects, omissibility in control and coordination constructions, constraints on linear ordering, etc., etc. (compare Bickel 2007: 242). The better, i.e., the more fine-grained and better operationalized these techniques and concepts, the

better both description and typology. Hence, work on these techniques and concepts constitute a (largish!) common ground between the two linguistic subdisciplines, and category controversies such as the continued debate regarding the applicability and usefulness of the concept "subject" in crosslinguistic comparison have been enormously fruitful and productive in bringing to the fore these more fine-grained and crosslinguistically applicable concepts. Thus, it is probably more correct to conceive of this common ground between typology and language documentation and description as one of mutual concern rather than as a unidirectional provision of "services" from typology to the latter.

I will conclude with a look at two further areas of mutual concern for the two subdisciplines.

3 Areas of common concern – where language documentation would like to learn more from typology in future

As indicated above, translation of discourse segments of many types and genres, from conversation to oral poetry, is a central task and activity in language documentation. Without it, the materials recorded for a language documentation would remain inaccessible to everyone but speakers of the language. Typically, the language used for translation is either a national language or one of academic communication, English and Spanish being by far the most frequent choices these days. Ideally, a good translation would be able to capture the basic flavor as well as the fine nuances of the original, i.e., the specific way linguistic meaning is construed in the source language.

Typology faces the same challenge when engaging with lexical and constructional meaning, a challenge it has yet to address in a principled and systematic way. Humboldt, Sapir, and many generations of typologists who followed them have occasionally made use of the ad hoc device of so-called literal translation, i.e., seeking ways of representing the meaning of a construction by using a frequently stilted and unidiomatic expression in the target language. Thus, for example, Humboldt (1838: 351) notes when discussing Tagalog voice alternations that the Tagalog utterance (in modern standard orthography) *bigyan mo ako ng tubig* [give.LOC.VOICE 2SG.GEN 1SG GEN water] means 'der wahren Construction nach: Gebungsort dein (sei) ich des Wassers, oder ich sei der Ort, wo du das Wasser hingiebst' ('according to its true construction: giving-place yours (be) I of the water, or I be the place where you give

the water to'). This practice is no longer common in mainstream typology, in part rightly so considering its unsystematic and ad hoc nature. However, there is also the much more problematic tendency simply to gloss over the issue of the crosslinguistic variability of meanings – other than the most obvious instances on the lexical level – and proceed on the assumption that the meanings of the items and constructions compared are at least roughly equal. In this regard, then, typology and language documentation are in need of better diagnostics for meaning differences and of practical means for representing such differences. The crucial task is to be able to separate culturally specific implicatures and presuppositions from possibly universal meaning components and to represent the former in a transparent way.

Turning finally to a completely unrelated point of common concern for typology and documentation, a major unresolved challenge for both is to find efficient and sustainable ways of communicating the nature and importance of linguistic diversity beyond the narrow confines of the discipline, especially also to non-academics. In the context of language documentation, this issue arises in particular in endangered speech communities who struggle to maintain their linguistic (and cultural) identity. While the own idiom is often valued highly, there is also a widespread tendency for purism and hence a lack of understanding the inherent variability of languages, which in the long run tends to be detrimental to language maintenance. On a very different level, such communities are usually embedded in and dependent on political and economic environments which either lack an appreciation for, or are actively hostile to, linguistic and cultural diversity. In such contexts, properly localized arguments in favor of linguistic diversity that have the potential to impact on public debate and opinion are needed.

For typology, the need to be able to communicate its basic concerns and results to the wider public may not be immediately obvious. However, at a time when massive migrations affect everyday life in most parts of the world, it seems obvious that typologists should be able to contribute to the often heated debates surrounding issues of diversity. Perhaps most importantly, schools are in need of materials for explaining and cherishing linguistic diversity. I have recently heard educators discussing the development of examples in artificial languages to illustrate the variability of grammatical structures (word order, for example). Typologists could do a much better job in this regard as the diversity actually attested in real languages provides a wealth of useful and instructive examples. They probably should team up with educators to get the didactics right. See http://de.languagesindanger.eu/ for an example.

Abbreviations: 1/2 = 1st/2nd person; GEN = genitive; LOC = locative; SG = singular.

References

Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11. 239–251.

Bond, Oliver. 2010. Language documentation and language typology. In Peter K. Austin (ed.) *Language documentation and description*, Vol. 7, 238–261. London: SOAS. http://www.elpublishing.org/PID/087

Croft, William, 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Cysouw, Michael & Bernhard Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. *Sprachtypologie und Universalienforschung* 60. 95–99.

Dingemanse, Mark, Francisco Torreira & N. J. Enfield. 2013. Is "huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS ONE* 8(11). e78273. http://doi:10.1371/journal.pone.0078273

Enfield, N. J., Tanya Stivers & Stephen C. Levinson. 2010. Question-response sequences in conversation across ten languages: An introduction. *Journal of Pragmatics* 42. 2615–2619.

Epps, Patience. 2010. Linguistic typology and language documentation. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 634–649. Oxford: Oxford University Press.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86. 663–687.

Himmelmann, Nikolaus P. 2012. Linguistic data types and the interface between language documentation and description. *Language Documentation & Conservation* 6. 187–207.

Humboldt, Wilhelm von. 1838. *Über die Kawi-Sprache auf der Insel Java*. Vol. 2. Berlin: Druckerei der königlichen Akademie der Wissenschaften.

Lazard, Gilbert. 2002. Transitivity revisited as an example of a more strict approach in typological research. *Folia Linguistica* 36. 141–190.