

## Meeting the transcription challenge

Nikolaus P. Himmelmann  
*Universität zu Köln*

The major challenge for language documentation in the next decade or two is what could be called the *transcription challenge*. This is a multilayered challenge that goes far beyond the practical challenge of speeding up the transcription process. Transcription, as practiced in language documentation, involves language making and changes the language ecology. Despite its centrality to language documentation, transcription remains critically undertheorized and understudied. Further progress in language documentation, and ultimately also its overall success, crucially depends on further investigating and understanding the transcription process, broadly conceived.

**1. Phonetic transcription, discourse transcription and the transcription bottleneck<sup>1</sup>** In linguistics, the term *transcription* is perhaps most closely associated with (narrow) **phonetic transcription**. The *International Phonetic Alphabet* (IPA) provides a means to capture core characteristics of the articulatory movements involved in speaking and thus to represent spoken language in writing. The transcription challenge discussed here does not pertain to this practice, as most challenges for phonetic transcription have successfully been resolved over the last century and a half, at least with regard to the segmental level. As a background to what follows, however, it would do well to remember that developing the IPA required solving major conceptual and practical issues. Perhaps more importantly, it would do well to remember that the representation of the suprasegmental aspects of speech continues to be a challenge. See Ladd (2014) for pertinent discussion.

Another level of transcription pertains to the creation of a written representation of recordings of more or less natural communicative events (everyday conversations, narratives, interactive games, speeches, etc.). The written representation typically provides the basis for the further analysis of such events. Transcription practices in this domain have been the object of theoretical and practical reflections in (different traditions

<sup>1</sup>I am very grateful to two anonymous reviewers and editor Bradley McDonnell. In my experience, it has been rare to get such perceptive and helpful input in the reviewing process. Thank you very much! Thank you very much also to Katherine Walker for thoroughly editing English grammar and style.

of) discourse and conversation analysis, Elinor Ochs' (1979) paper *Transcription as Theory* being the classic example (cp. Edwards & Lampert 1993 for a collection of papers on this topic and Bucholtz 2007 for a more extensive bibliography). Unfortunately, neither the theoretical concerns nor the practical guidelines developed in these traditions (e.g. Du Bois et al. 1992, Selting et al. 2009) have had a major impact on practices in field linguistics and language documentation. That is, despite the fact that **discourse transcription** is at the core of documentary linguistic activity, it remains a topic that is rarely discussed in the field.<sup>2</sup> Consequently, there is little agreement about very fundamental decisions such as how to segment spoken language (cp. Himmelmann 2006 for a short overview of the main issues). More often than not, segmentation units above the word (i.e. prosodic units and/or syntactic phrases) are not explicitly discussed or justified, and are thus difficult to reconstruct and evaluate for users of a documentation. To make discourse transcription a major topic in the field, then, is one aspect of meeting the transcription challenge in language documentation.

A further, related aspect of this challenge is the **transcription bottleneck** (cp. Seifart et al. 2018). Given state-of-the-art recording technologies and a community supportive of creating a comprehensive record of their speaking practices, it is now relatively easy not only to compile a largish collection of documentary recordings, but also to archive them and make them available to other interested parties. But making them truly accessible by adding transcription and translation is a different matter altogether. Estimates of the factor involved here vary, depending on recording quality, the number of speakers involved, etc. Factors smaller than 10 (i.e. ten minutes are necessary to transcribe and translate one minute of recording) are rarely mentioned, and factors as high as 150 and higher are not unrealistic in the case of complex multiparty conversations. Some aspects of the transcription-cum-translation process are fairly mechanical and highly repetitive. Without doubt, support from machine-based speech processing could be of great help in speeding up the transcription process.

A number of efforts in this regard have been undertaken in recent years, as of yet without major success.<sup>3</sup> Given the relatively small amounts of data and manpower typically available in the case of underdocumented languages, it is clear that any success of such efforts will not even remotely approach the power of the automatic transcription tools for natural speech currently emerging for major national languages, in particular English. But even the automatization of tasks such as identifying different speakers (speaker diarization), proposing an initial rough, pause-based segmentation and recognizing high-frequency items and phrases would already be of major help in processing documentary recordings. An important side effect of these efforts, inasmuch as they are not confined to mere phone recognition, is the fact that they force the community to become more explicit about its transcription practices, thus addressing the issue of discourse transcription mentioned above.

<sup>2</sup>To my knowledge, Crowley (2007:137-141) is the only work on linguistic fieldwork that discusses the practicalities of transcribing larger amounts of narrative and conversational speech, going beyond the problems of basic procedure and properly capturing sound. In other subfields of linguistics, (introductory) discussions of transcription tend to be considerably more comprehensive and sophisticated. See Nagy & Sharma (2013) for an example.

<sup>3</sup>Including the *Transcription Acceleration Project* at the Australian Centre of Excellence for the *Dynamics of Language* (<http://www.dynamicsoflanguage.edu.au/news-and-media/latest-headlines/article/?id=early-results-from-survey-exploring-transcription-processes>) and the *Kölner Zentrum Analyse und Archivierung von Audiovisuellen Daten* (<http://ifl.phil-fak.uni-koeln.de/32830.html>).

**2. The real challenge: understanding the transcription process** However, the points mentioned so far are only the beginnings of what I consider to be the core of the transcription challenge: reaching a **better understanding of the transcription process** itself and its relevance for linguistic theory. The two central questions here are:

- (1) What do speakers and researchers actually do when they transcribe (or assist in transcribing)?
- (2) How does the fact that transcription converts specimens of spoken language into a written representation—possibly accompanied by further annotations such as a translation and notes on grammar and cultural background—affect the overall language ecology in the community providing the specimens of natural speech compiled in a language documentation?

As to the first question, it would be rather naïve to consider transcription exclusively, or even primarily, a process of mechanically converting a dynamic acoustic signal into a static graphic/visual one. Transcription involves interpretation and hence considerable enrichment of the acoustic signal. That is, transcription necessarily involves hypotheses as to the meaning of the segment being transcribed and the linguistic forms being used. Linguistic forms tend to be underdetermined by the acoustic signal, as everyone who has ever engaged in transcribing spontaneous speech knows (missing or unclear segments, ambiguous reduced forms, etc.; see also Hermes & Engman (2017: 65 *passim*)). But how exactly does interpretation and enrichment actually work in the transcription process?

In Jung & Himmelmann (2011), we provide some preliminary observations concerning the transcription process, based on recorded transcription sessions where a linguist works together with a native speaker. Typical reactions to the transcription task, bearing witness to the creative aspects of the transcription process, include the tendency for the native speaker to paraphrase what is said rather than repeating it more or less directly. There is also the tendency to edit out elements typical of spoken language such as particles, hesitations and the like. The converse tendency is to edit in material that is deemed to make the transcript ‘better’, ‘more correct’ or ‘clearer’, such as using fuller verb forms, pronouns, and so on. Similar changes occur when transcripts are further edited for publication (Mosel 2014). Marten & Petzell (2016) give a very instructive example of the kinds of ‘purifications’ that often occur in multilingual settings with a major dominant language (in their case study Swahili).

Furthermore, in most instances, transcription also involves language learning. In fact, as illustrated and further discussed in Hermes & Engman (2017), transcription may be used as one way to learn a language, especially when younger speakers collaborate with older speakers in a transcription task. As a consequence, there are overlaps between the kinds of processes that typically occur in (adult) second language acquisition and transcription, and hence there is major potential for cross-fertilization between these two fields of linguistic inquiry.

In order to better understand the transcription process and thus to supply an answer to question 1 above, these and other aspects of the transcription process need to be studied much more systematically. Given the fact that written transcripts are underdetermined by the acoustic signal, transcriptions of the same recording produced by different transcribers/transcription teams will differ in some details. Consequently, questions such as the following arise: How variable are the transcriptions typically produced in language documentation, i.e. to what extent are they underdetermined by the acoustic signal (which of course itself will vary across recordings)? What is the potential impact of this variability for subsequent analyses? What kinds of phenomena are particularly salient for transcribers and how do these intersect with the phenomena salient for second language learners? In order to answer these questions, we need transcription experiments that target those features of transcripts that are particularly prone to variation and those that tend to prompt special attention. Bucholtz (2007), for example, investigates variation in format choices, orthographic variation and variation in translation. Himmelmann et al. (2018) is another example that investigates a feature well-known to be highly variable, namely prosodic segmentation.

However, while better understanding sources of variability and variation in transcription is important for putting language documentation on a safer methodological footing, something else may potentially be of even greater significance, not only for language documentation but for the language and cognitive sciences more generally. When speakers edit in and edit out, what kind of knowledge and norms do they base their decisions on? Why is it that all over the world, speakers of very different languages, living in very diverse linguistic settings, have clear ideas about the fact that some parts of a recorded spontaneous utterance are ‘not relevant’ and hence should be edited out? And that the acoustically observable form X is the short/reduced variant of form Y? And that expression A is ‘better’ and ‘more complete’ than expression B?

The first, seemingly trivial, answer that springs to mind is that all of these ideas and reactions are based on a written standard and prescriptive traditions learned in schools. But then, what about linguistic varieties without a written standard and not used in (formal) education? Of course, very few, if any, settings exist nowadays where there are not at least a few speakers of a given variety who are also familiar with a written language standard and have received some formal schooling. Hence, it is possible that ideas as to how written language should look and what the proper ways of speaking are have disseminated from written language traditions and schooling. But does this really suffice to explain that speakers who are illiterate and do not know a regional or national standard(ized) language show the same tendencies for editing in and out when assisting with transcription?

An alternative explanation for the typical reactions shown by speakers when assisting in a transcription task is that transcription taps into a form of linguistic knowledge that differs from the linguistic knowledge underlying linguistic behavior in spontaneous interactions. Transcription is of course not part of anyone’s native linguistic repertoire—it is a new way of dealing with language for those who engage in it for the first time. However, while being new in its specifics, it probably belongs to a larger class of activities that involve **metalinguistic knowledge and awareness**, i.e. a reflective mode (as opposed to a production mode) in dealing with language. Other examples where this mode comes into play are language games, verbal arts, adult language acquisition and conscious choices made in multilingual settings. If metalinguistic knowledge and awareness are factors in transcription, then, obviously, investigations of the transcription process should

be informed by research on metalinguistic knowledge and awareness in other domains.<sup>4</sup> At the same time, transcription opens a new venue for researching this type of linguistic knowledge.

**3. Productive and reflective modes of language use** It is well known that there are fundamental differences between speaking and writing, with regard to both the linguistic structures being used and the cognitive resources that are deployed in the respective production processes (e.g. Chafe 1982, 1994; Akinnaso 1982, 1985; Biber 1988, 2014). The hypothesis proposed here is that these differences are not exclusively bound to the differences in the communicative channel (auditory vs. visual) but also occur in the oral language domain itself, as also assumed in Labov's (1972) notion *attention to speech* and in much contemporary work on 'style' (e.g., Coupland 2007).<sup>5</sup> That is, there are language-related activities, including language games, verbal arts and transcription, that make use of a reflective mode in dealing with (spoken) language, which is different from the mode employed in producing spontaneous speech. This hypothesis and the considerable body of work on the differences between speaking and writing, then, provide the basic framework for investigating what speakers actually do when transcribing, and what this implies for linguistic and, more generally, cognitive theories.

A demanding, but also exciting, research program follows when transcription is approached in this way. Transcription practices need to be documented more systematically in order to get the full picture of which strategies and which forms of knowledge are applied in transcription.<sup>6</sup> Other linguistic activities which tap into metalinguistic knowledge and awareness need to be identified to allow us to develop a comprehensive view of the reflective use of language. As metalinguistic knowledge and awareness are not accessible to direct observation, many methodological challenges have to be overcome in determining how insights about this knowledge type can be derived from observable linguistic behavior (such as the behavior shown by native speakers in transcription tasks) and tested in well-designed experiments.

**4. Transcription as language making** Inasmuch as transcription necessitates the development of a new way of representing speech, it is a prototypical instance of language making.<sup>7</sup> This aspect of transcription is often underestimated, because many practitioners tend to underestimate, and hence fail to take into account, the amount of interpretation and enrichment involved in the transcription process as highlighted in the preceding two sections. But, in addition to the cognitive aspects discussed above, when its language-making potential is taken seriously there is also a social aspect to transcription.

<sup>4</sup>The journal *Language Awareness* provides examples, which, however, mostly concern the North American and European settings and frequently pertain to the role of awareness in language learning and teaching. See Verschik (2015) for a brief survey of the role of language awareness in multilingual settings and language contact.

<sup>5</sup>There is an almost forgotten tradition dating back to the 19<sup>th</sup> century that has developed this hypothesis, using the term *Ausbau* for what here is called *reflective mode*. Maas (2009, 2010) provides details and arguments.

<sup>6</sup>In fact, as one reviewer rightly points out, current practice is often wanting even on the most basic level of including the information of who was involved in the transcription (only a member of the community? the researcher? both?).

<sup>7</sup>Note that even in communities with well-established writing traditions both spontaneous and scientific representations of spoken language usually involve the invention of new conventions such as new uses of punctuation and various attempts to capture salient aspects of speech by indicating lengthening, stress and frequent fast-speech forms (e.g., *gonna*), etc.

Transcription introduces a new element into the linguistic repertoire of those who engage with it. What is completely unclear to date is whether and to what extent this change may have repercussions for the linguistic practices of the community at large. At first sight, potential repercussions may appear to be negligible. Typically, only one or two members of the community work on transcription in close collaboration with the documentation team. The rest of the community would appear not to be affected by it and is usually also not very interested in what is for them an obscure activity. However, we do not know for sure that this is indeed the case. To date, what the community at large actually knows about the transcription process and whether transcription can influence the overall **language ecology**, however subtly, has never been investigated.


In this regard, it would do well to remember that the production of written language materials—often a dictionary or a reading primer—obviously changes the language ecology in all those instances where the linguistic variant in question was not represented in writing before the documentation project started. The production of such materials involves standardization on many levels: determining base forms for lexical entries, orthographic conventions (e.g., how clitics are written), and so on. Many of these decisions are also part of the transcription process. Usually no attempt is made to include a larger group of community members in these decisions at the transcription stage. But it stands to reason that in one way or another they will indeed affect the community, given that it is a basic goal of modern language documentation to be available and accessible to the community. Hermes & Engman (2017) provide a very instructive example of how the inclusion of a larger group of speakers in transcription changes the documentation process, making it more accessible and relevant for the speech community (in this case, for revitalization).

**5. Conclusion** It is only a minor exaggeration to say that language documentation is all about transcription. The major argument for a separation of documentation from description—on the conceptual level, not in actual practice where such a separation is impossible—has always been that such a separation helps to focus on those aspects of linguistic fieldwork and language description that tended to be overlooked in descriptive linguistics as practiced throughout the 20<sup>th</sup> century. Transcription is at the very center of these overlooked and underscrutinized aspects, and documentary linguistics has not yet properly engaged with this core challenge in its subject matter. The brief remarks offered here hopefully make clear not only the importance of properly investigating the transcription process, but also the potential and promises of such investigations for understanding language and the human mind.

## References

- Akinnaso, F. Niyi. 1982. On the differences between spoken and written language. *Language and Speech* 25. 97–125.
- Akinnaso, F. Niyi. 1985. On the similarities between spoken and written language. *Language and Speech* 28. 323–359
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas. 2014. Using multidimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast* 14. 7–34.
- Bucholtz, Mary. 2007. Variation in transcription. *Discourse Studies* 9. 784–808.
- Chafe, Wallace. 1982. Integration and involvement in speaking, writing, and oral literature. In Deborah Tannen (ed.), *Spoken and written language: Exploring orality and literacy*, 35–53. Norwood: Ablex.
- Chafe, Wallace. 1994. *Discourse, consciousness, and time*. Chicago: University of Chicago Press.
- Coupland, Nikolas. 2007. *Style: Language variation and identity*. Cambridge: Cambridge University Press.
- Crowley, Terry. 2007. *Field linguistics: A beginner's guide*. Oxford: Oxford University Press.
- DuBois, John W., Stephan Schuetze-Coburn, Danae Paolino & Susanna Cumming. 1992. *Discourse transcription*. Santa Barbara Papers in Linguistics 4. Santa Barbara, CA: Department of Linguistics, University of California, Santa Barbara.
- Edwards, Jane & Martin D. Lampert (eds). 1993. *Talking data: Transcription and coding in discourse research*. Hillsdale: Lawrence Erlbaum.
- Hermes, Mary & Mel M. Engman. 2017. Resounding the clarion call: Indigenous language learners and documentation. In Wesley Y. Leonard & Haley De Korne (eds.), *Language Documentation and Description*, vol 14. London: EL Publishing, 59–87. <http://www.elpublishing.org/PID/152>
- Himmelmann, Nikolaus P. 2006. The challenges of segmenting spoken language. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 253–274. Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases in spontaneous speech – a cross-linguistic interrater study. *Phonology* 35. 207–245.
- Jung, Dagmar & Nikolaus P. Himmelmann. 2011. Retelling data: Working on transcription. In Geoffrey Haig, Nicole Nau, Stefan Schnell & Claudia Wegener (eds.), *Documenting endangered languages*, 201–220. Berlin: Mouton de Gruyter.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: The University of Pennsylvania Press.
- Ladd, D. Robert. 2014. *Simultaneous structure in phonology*. Oxford: Oxford University Press.
- Maas, Utz. 2009. Orality vs. literacy as a dimension of complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 164–177. Oxford: Oxford University Press.
- Maas, Utz. 2010. Literat und orat. Grundbegriffe der Analyse geschriebener und gesprochener Sprache. *Grazer Linguistische Studien* 73. 21–150. <http://unipub.uni-graz.at/gls/periodical/pageview/1276544>

- Marten, Lutz & Malin Petzell. 2016. Linguistic variation and the dynamics of language documentation: Editing in 'pure' Kagulu. In Mandana Seyfeddinipur (ed.), *African language documentation: New data, methods and approaches* (Language Documentation & Conservation Special Publication 10), 105–129. <http://hdl.handle.net/10125/24651>
- Mosel, Ulrike. 2014. Putting oral narratives into writing – experiences from a language documentation project in Bouganville, Papua New Guinea. In Bernard Comrie & Lucía Golluscio (eds.), *Language contact and documentation / Contacto lingüístico y documentación*, 321–342. Berlin: De Gruyter.
- Nagy, Naomi & Devyani Sharma. 2013. Transcription. In Robert J. Podesva & D. Devyani Sharma (eds.), *Research Methods in Linguistics*, 235–256. Cambridge: Cambridge University Press.
- Ochs, Elinor. 1979. Transcription as theory. In Elinor Ochs & Bambi B. Schieffelin (eds.), *Developmental pragmatics*, 43–72. New York: Academic Press.
- Seifart, Frank, Nicholas Evans, Harald Hammarström & Stephen C. Levinson. 2018. Language documentation 25 years on. *Language* 94, e324-e345. DOI: 10.1353/lan.2018.0070
- Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg R. Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzluft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock & Susanne Uhmann. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 10. 353–402. <http://www.gespraechsforschung-ozs.de>.
- Verschik, Anna. 2015. Language contact, language awareness, and multilingualism. In Jasone Cenoz, Durk Gorter & Stephen May (eds.), *Encyclopedia of language and education*. Dordrecht: Springer. doi: 10.1007/978-3-319-02325-0\_21-1.

Nikolaus P. Himmelmann  
sprachwissenschaft@uni-koeln.de  
 [orcid.org/0000-0002-4385-8395](https://orcid.org/0000-0002-4385-8395)