

# PERAN TEKNOLOGI DAN KERJA SAMA TIM DALAM PENDOKUMENTASIAN BAHASA

*On the Role of Technology and Team Work in Language Documentation*

**Nikolaus P. Himmelmann**

Universität zu Köln

Pos-el: sprachwissenschaft@uni-koeln.de

## **Abstrak**

Jika dibandingkan dengan pendeskripsian bahasa dengan pendekatan yang lebih tradisional, ada dua ciri penting pendokumentasian bahasa, yaitu perhatian eksplisit terhadap teknologi yang dapat membuat pekerjaan pendokumentasian bahasa yang belum terdokumentasi dengan baik menjadi lebih efisien dan penekanan tentang perlunya tim yang terdiri atas orang-orang dengan keahlian yang berbeda-beda dan jenis keahlian berbeda untuk bekerja bersama dalam usaha pendokumentasian dan konservasi. Jika dilihat dari kontribusinya saat ini, kedua hal tersebut berkaitan erat satu sama lain dan saling mendukung satu sama lain. Repositori-bahasa daring (arsip) memungkinkan kelompok pengguna yang berbeda untuk mengakses dan bekerja dengan materi yang dikumpulkan pada pendokumentasian bahasa. Pada saat yang bersamaan, arsip ini memungkinkan pengguna untuk memperbarui dokumentasi dengan cara menambahkan data baru atau membuat anotasi dan analisis lanjutan. Sejalan dengan itu, untuk mendapatkan dokumentasi yang baik, menjadikan penutur asli dan komunitas tutur sebagai partner aktif dalam proyek dokumentasi merupakan hal yang esensial. Hal itu bahkan lebih esensial bagi upaya revitalisasi. Dalam hal ini, teknologi berbasis telepon pintar bisa membantu memperluas kemungkinan partisipasi penutur asli dan pihak lain yang tertarik.

**Kata-kata kunci:** arsip bahasa, repositori dinamis, aplikasi telepon pintar, kerja sama tim

## **Abstract**

*Two important features of language documentation, as opposed to more traditional approaches to language description, are its explicit concern for technologies that help to make work on underdocumented languages more efficient and its emphasis on the need for teams of people with different skills and kinds of expertise to work together in documentation and conservation efforts. As shown in the current contribution, these two concerns are closely interlinked and mutually reinforce each other. Online dynamic language repositories ('archives') make it possible for different user groups to access and work with the materials compiled in a language documentation. At the same time, they make it possible for such users to enhance the documentation by adding new data or further annotation and analysis. Similarly, for good documentations it is essential that native speakers and the speech community at large become active partners in the documentation enterprise. This holds even more so for revitalization efforts. Here smart-phone based technology may help to widen the participatory possibilities for native speakers and other interested parties.*

**Keywords:** language archives, dynamic repository, smart phone applications, team work

## PENDAHULUAN

Mengikuti Himmelmann (2006:1) tujuan pendokumentasian bahasa dapat didefinisikan sebagai penciptaan “rekaman bahasa yang kekal dan multiguna”, di sini ‘bahasa’ harus dipahami dalam pengertian luas yang meliputi semua repertoar komunikasi yang tersedia di komunitas tutur pada suatu masa. Inti pendokumentasian bahasa terdiri atas contoh perilaku linguistik (idealnya rekaman video tentang interaksi linguistik) dan contoh pengetahuan metalinguistik (komentar penutur asli dan pandangan tentang bentuk dan praktik linguistik, sering kali berdasarkan elisitasi atau wawancara, dan juga diamati pada aktivitas inti dokumentasi, yaitu pentranskripsian dan penerjemahan).

Agar dapat digunakan dengan mudah oleh pihak lain yang tertarik, khususnya mereka yang tidak terlibat langsung dalam usaha pendokumentasian bahasa, dokumentasi bahasa harus dikurasi secara digital. Seksi 2 menjelaskan apa itu kurasi digital dengan berfokus pada arsip bahasa. Aplikasi telepon pintar menyediakan tempat baru untuk mengakses dan berkontribusi untuk pendokumentasian bahasa. Topik ini dibahas dengan ringkas pada Seksi 3 yang berfokus pada dunia kolaborasi dalam pendokumentasian bahasa. Potensi maksimal dari pendokumentasian bahasa dapat tercapai ketika orang yang memiliki kemampuan dan jenis keahlian yang berbeda bekerja secara bersama-sama.

### 1. Kurasi Digital dan ‘Arsip Partisipatori’

Dalam linguistik dokumentasi, mengacu repositori digital untuk materi dokumentasi sebagai arsip bahasa merupakan hal yang umum. Bandingkan contoh (lihat Berez-Kroeker & Henke (2018:354ff) untuk contoh yang lebih banyak dan lebih rinci):

- *The Language Archive* (<https://tla.mpi.nl>), yang meliputi *DoBeS Archive* (<https://tla.mpi.nl/resource/dobes-archive/>);
- *the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)*, (<http://www.paradisec.org.au/>);
- *the Endangered Languages Archive (ELAR)*, (<https://elar.soas.ac.uk/>);
- *the Archive of the Indigenous Languages of Latin America (AILLA)*, (<https://ailla.utexas.org/>).

Pemanfaatan arsip dalam pengacuan repositori semacam itu telah mengakibatkan kebingungan tentang penggunaan istilah arsip yang diasosiasikan dengan entitas tertutup dan statis. Secara tradisional, arsip (kertas) hanya terdiri atas koleksi tertutup, seperti tulisan (huruf, draf manuskrip, dan lain-lain) tentang orang penting atau rekaman tentang lembaga administrasi atau politik pada suatu masa (contohnya, debat di Parlemen Inggris pada 1957). Dokumentasi bahasa, idealnya, meliputi repositori yang terbuka untuk memasukkan data dan anotasi baru. Seperti yang akan kita lihat selanjutnya, konsep baru tentang *Arsip Partisipatori* tampaknya akan lebih cocok dengan pendokumentasian bahasa. Sebelum mengklarifikasi isu ini lebih lanjut, penentuan jenis data yang dapat dimasukkan dalam repositori menjadi hal penting.

Seperti disebutkan sebelumnya, dokumentasi bahasa terdiri atas, di satu sisi, contoh perilaku linguistik yang teramati, yaitu rekaman audio-video dengan materi pendukung yang dapat membantu pemahaman dan penginterpretasian contoh yang ada.<sup>1</sup> Di sisi lain, arsip terdiri atas contoh pengetahuan metalinguistik yang juga bisa berbentuk rekaman (seperti rekaman seorang penutur dan linguist mendiskusikan transkripsi atau transliterasi; atau rekaman tentang seorang penutur yang sedang menjelaskan taksonomi tumbuhan atau hubungan kekeluargaan dengan menggunakan bahasa kontak). Akan tetapi, arsip ini sering kali hanya terdiri atas catatan yang ditulis linguist saat berdiskusi dengan seorang penutur tentang bentuk dan praktik linguistik. Dalam hal ini, bahan pendukung harus ditambahkan.

Materi pendukung, atau dalam istilah yang lebih teknis *apparatus*, meliputi transkripsi dan transliterasi peristiwa tutur yang direkam, anotasi lanjutan tentang aspek tata bahasa dan budaya dari peristiwa tutur, foto-foto yang mendokumentasikan peristiwa tutur, catatan tangan yang dibuat saat perekaman, dan materi lain yang relevan untuk memahami dan menginterpretasi rekaman.

Sementara itu, pada prinsipnya ada beberapa cara untuk mengorganisasi data yang dikumpulkan untuk dokumentasi bahasa: unit dasar pengorganisasian data yang ditemukan paling bermanfaat adalah *sesi*, yaitu sepaket fail yang memuat satu peristiwa komunikasi seperti yang terlihat di dalam daftar sesi. Paket sesi biasanya terdiri atas fail

---

<sup>1</sup> Untuk membuat eksposisi ini tetap sederhana dan untuk tujuan diskusi ini, saya berasumsi bahwa dokumentasi bahasa modern sebagian besar berbentuk digital. Rekaman audio-video dan hampir semua jenis materi pendukung sudah diproduksi secara digital, pengecualian untuk catatan tertulis yang biasanya harus dipindai agar dapat dimasukkan dalam repositori digital. Pemuatan materi peraturan, seperti rekaman analog dari masa lampau, foto lama, dan catatan lapangan yang dapat memunculkan permasalahan tidak didiskusikan di sini.

dengan tipe yang berbeda-beda, seperti fail media dengan format-format yang berbeda (wav, mpeg, jpg), fail XML untuk anotasi, pdf untuk catatan, dan dokumen tertulis lainnya yang juga berperan dalam peristiwa tutur.

Durasi peristiwa tutur yang dikehendaki bisa pendek (misalnya, gurauan atau cerita pendek) atau sangat panjang (misalnya, pertemuan politik yang berlangsung selama beberapa hari), ukuran dan kompleksitas sesi bervariasi bergantung pada peristiwa tutur yang direkam. Peristiwa tutur juga bisa berbeda berdasarkan tipe interaksinya: percakapan sehari-hari, percakapan yang berorientasi pada tugas, narasi, elisitasi, dll.

Setiap sesi harus berisi fail metadata yang tidak hanya merinci fail-fail yang ada dalam satu sesi, tetapi juga memuat informasi dasar tentang kapan, di mana, siapa, dan apa yang terlibat dalam suatu peristiwa tutur. Kapan peristiwa tutur itu terjadi? Siapa yang berpartisipasi? Apa topik dasar dan tujuannya? Peralatan apa yang dipakai untuk merekam? Sesi itu termasuk proyek dokumentasi yang mana?

Koleksi dari keseluruhan sesi, yaitu seluruh dokumentasi bahasa, juga memerlukan metadata, yang menyatakan tujuan proyek dokumentasi, siapa yang berpartisipasi, konvensi apa yang diikuti (ortografi, penglosan, dll.) dan informasi lain yang diperlukan untuk menginterpretasi data yang ada dalam koleksi. Bagian utama dari metadata pada level koleksi adalah pengantar tentang bahasa dan budaya dari komunitas yang didokumentasikan. Di sini perbedaan antara dokumentasi dan deskripsi bahasa dari segi analisis yang lebih mendalam menjadi kabur. Semakin analitis materi yang terdapat pada aparatus di level koleksi (misalnya, tata bahasa lengkap, bukan sekadar tata bahasa ringkas; kamus lengkap dengan struktur subentri yang detail, bukan sekadar *database* lembar kerja yang berisi entri leksikal, dll.), semakin bermanfaat koleksi tersebut untuk memahami dan menginterpretasi data yang terkumpul pada level sesi.

Tabel 1 (dari Himmelmann 2016:14) menyimpulkan struktur dokumentasi bahasa sebagai berikut.

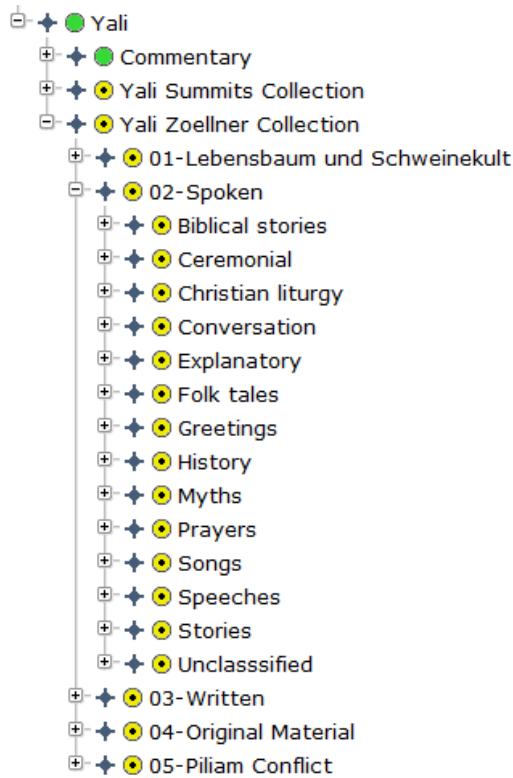
Tabel 1. Format dasar dokumentasi bahasa

Perekaman terhadap perilaku linguistik yang teramati dan pengetahuan metalinguistik	Per Sesi	Untuk Seluruh Dokumentasi
	Metadata Waktu dan lokasi	Metadata Lokasi komunitas yang

	perekaman Partisipan Tim perekaman Peralatan perekaman Deskripsi isi ... Anotasi Transkripsi Transliterasi Glos lanjutan dari segi linguistik dan etnografi serta komentar	didokumentasikan Tim proyek yang berkontribusi pada dokumentasi Partisipan dalam dokumentasi Ucapan terima kasih ... Sumber akses umum Pendahuluan Konvensi ortografi Etnografi ringkas Tata bahasa ringkas Konvensi pengglosan Indeks Tautan ke sumber lain ...
--	---	---

Sepengetahuan saya, semua repositori untuk dokumentasi bahasa yang ada saat ini pada prinsipnya sepakat untuk membuat sesi sebagai inti dari struktur repositori; isu yang masih terbuka adalah tentang apakah struktur pohon yang lebih terperinci diperlukan atau tidak, yaitu struktur yang menempatkan sesi dengan isi yang sama pada cabang yang sama. Pada pohon semacam ini, sesi monolog dapat dibedakan dengan sesi polilog. Dalam sesi monolog, seseorang dapat membedakan cerita rakyat, cerita kehidupan pribadi, teks procedural, dll., seperti yang ditunjukkan oleh pohon bagian dari arsip DoBes yang lama pada Gambar 1. Alternatif untuk struktur pohon semacam itu bersifat *faceted browsing* (multiramban). Akan tetapi, struktur ini menuntut metadata yang sistematis untuk setiap sesi dan dipercaya mengandung informasi yang diperlukan untuk setiap kategori perambanan.

Gambar 1: contoh struktur pohondokumentasi bahasa (versi arsip DoBes yang lama)



Dari daftar elemen yang harus dimasukkan dalam dokumentasi bahasa, jelas bahwa kurasi digital merupakan suatu keharusan dalam dokumentasi bahasa. Tentu saja dengan prinsip bahwa seseorang dapat menyimpan kaset rekaman dan transkripsi cetak, terjemahan, dll. secara bersamaan. Namun, untuk menggunakan materi yang tersimpan dengan cara ini akan sangat rumit dan terbatas hanya untuk orang yang memiliki akses ke tempat penyimpanannya. Ketersediaan rekaman video dan audio digital secara fundamental telah mengubah cara peneliti dan pihak lain yang tertarik untuk mengakses dan bekerja dengan tipe media ini. Kenyataannya, media-media itu sesungguhnya menjadi titik tolak model dasar sesi inti pada dokumentasi bahasa, yaitu transkripsi, terjemahan, dan anotasi lain yang idealnya selaras-waktu dengan rekaman. Hal itu tidak hanya akan memudahkan pengecekan rekaman yang asli, tetapi juga menjadi spesifik dan tepat untuk menunjukkan asal suatu anotasi. Dengan program seperti ELAN (<https://tla.mpi.nl/tools/tla-tools/elan/>), anotasi selaras-waktu dapat ditunjukkan dengan nyaman dan efisien. Program ini memiliki manfaat tambahan, yaitu dapat menyimpan anotasi dalam format XML tanpa hak milik yang sangat penting dalam penyimpanan yang bersifat abadi dan portabel (bandingkan Bird & Simons (2003) untuk informasi tentang pentingnya portabilitas dalam dokumentasi bahasa).

Kurasi digital tidak hanya penting untuk membuat dokumentasi bahasa dapat diakses oleh pihak lain yang tertarik, tetapi hal ini juga penting bagi anggota tim dokumentasi bahasa yang bekerja untuk menganotasi dan menganalisis data. Koleksi digital yang dikurasi dengan tepat juga merupakan lingkungan penelitian digital tempat peneliti mengakses dan membagi data. Karena semua fail media, khususnya video, yang cenderung berukuran besar, dokumentasi bahasa biasanya memperkecil ukuran media tersebut beberapa ratus *gigabyte* atau *terabyte*, bahkan lebih menjadi data yang ukurannya bisa ditangani dalam laptop standar atau komputer *desktop*. Maka dari itu, kurasi digital yang tepat menjadi urusan pribadi peneliti, karena jika tidak dilakukan, mereka tidak dapat bekerja secara produktif dengan data yang dikumpulkan dalam pendokumentasian bahasa.

Akhirnya, seperti yang sudah disebutkan sebelumnya, kurasi digital penting untuk bekerja secara kolaboratif pada kumpulan data yang telah ditentukan. Peneliti-peneliti berbeda yang bekerja bersama dalam sebuah tim dokumentasi perlu mengoordinasikan anotasi dan analisis mereka. Contoh yang tipikal adalah pangkalan data leksikal yang biasanya menjadi unsur utama dalam dokumentasi. Peneliti yang berbeda akan menambahkan entri ketika mengerjakan transkripsi atau mengelisisasi data dari penutur asli. Transkripsi dan terjemahan harus dicek dan direvisi, dan seterusnya. Semua proses ini akan berjalan dengan sangat baik jika dokumentasi tersedia dalam bentuk digital.

Yang paling penting adalah bahwa kelompok orang yang bekerja untuk memperkaya dan memperbesar dokumentasi bahasa seharusnya atau idealnya tidak terbatas pada anggota tim dokumentasi yang telah ditentukan dari awal. Akan tetapi, peneliti lain dan anggota masyarakat yang tertarik, khususnya mereka yang berasal dari komunitas tutur bahasa yang didokumentasikan harus memiliki kesempatan untuk mengomentari, menambah, dan memberi saran terhadap dokumentasi bahasa yang sedang dikembangkan dengan kesempatan untuk dapat mengakses dokumentasi dan memberikan komentar, perbaikan, dan saran pada aspek-aspek yang mereka kuasai. Contoh di luar ranah dokumentasi bahasa yang memungkinkan penulisan dan penerbitan yang bersifat kolaboratif disediakan oleh *Overleaf* atau *Google Docs*. Saat ini, repositori untuk dokumentasi bahasa belum menyediakan kesempatan kolaborasi yang sesungguhnya secara daring. Namun, hal itu sebenarnya merupakan langkah selanjutnya yang harus dicapai dalam pengembangan teknologi dokumentasi bahasa selanjutnya.

Kolaborasi pada paragraf berikut ini merupakan inti dari proposal terkini untuk merekonseptualisasi arsip sebagai Arsip Partisipatori (lihat Berez-Kroeker & Henke 2018, Seksi 3, untuk diskusi dan acuan yang lebih lengkap). Definisi sederhana tentang Arsip Partisipatori dikemukakan oleh Theimer (2011).

Sebuah organisasi, situs, atau koleksi tempat orang-orang selain pengarsip profesional mengontribusi pengetahuan atau sumber daya yang berakibat pada peningkatan pemahaman terhadap materi arsip, yang biasanya dalam lingkungan daring.

Perlu dicatat bahwa isu ini bukanlah isu khusus dalam hal dokumentasi bahasa, melainkan isu tentang semua aktivitas pengarsipan di semua ranah. Untuk repositori dokumentasi bahasa, Arsip Partisipatori tampak sebagai pendekatan yang tepat karena arsip ini menghindari bentuk arsip tradisional yang biasanya statis dan tertutup. Secara alamiah, arsip partisipatori membutuhkan keterbukaan dan pemutakhiran yang dinamis sehingga sesuai dengan tuntutan repositori dinamis untuk dokumentasi bahasa.

#### **Kolaborasi dan Aplikasi Mobil (*Mobile*)**

Melanjutkan diskusi tentang kolaborasi pada bagian akhir sesi sebelumnya, perkembangan terkini yang perlu disebutkan di sini adalah menggunakan aplikasi mobil untuk memperbanyak opsi pilihan partisipasi dalam dokumentasi bahasa, khususnya bagi penutur asli.

Dokumentasi bahasa yang baik memerlukan partisipasi dan kolaborasi dari penutur asli secara individu dan komunitas tutur secara keseluruhan. “Partisipasi dan kolaborasi” berarti bahwa anggota komunitas tutur berpartisipasi dalam pendokumentasian di luar perannya yang secara tradisional hanya diwawancara dan diamati, dengan kata lain, sebagai objek penelitian. Sebaliknya, praktik dokumentasi yang baik membutuhkan penutur asli untuk menjadi anggota tim dokumentasi yang tetap dan aktif, mengemukakan agenda, harapan, dan keinginan yang dapat diperhitungkan ketika merencanakan aktivitas proyek dokumentasi bahasa.

Dalam konteks ini, aplikasi mobil menyediakan cara untuk meniadakan partisipasi secara fisik di lokasi tertentu. Bird (2018:842) menjelaskan:

Penutur bahasa-bahasa yang hampir punah secara cepat dapat memperoleh jaringan internet yang luas melalui peralatan mobil. Sementara itu, teknologi mobil sosial melanjutkan untuk mentransformasi cara orang untuk bekerja bersama. Saya percaya bahwa kondisi tersebut tepat dengan pengembangan perangkat lunak (*software*) generasi baru untuk bahasa-bahasa hampir punah. Perangkat lunak ini akan memungkinkan cara baru bagi linguist untuk berkolaborasi dengan penutur di tanah



airnya sendiri dengan diaspora yang tersebar di seluruh dunia sehingga tercipta dokumentasi berskala besar yang berkualitas tinggi.

Seperti didiskusikan dalam artikel buku panduan Bird (2018) ada banyak isu yang didiskusikan dalam pengembangan aplikasi semacam itu. Yang paling penting, aplikasi tersebut hanya akan digunakan jika penutur asli terlibat dalam pengembangannya. Perkembangan terjauh sampai saat ini adalah aplikasi kamus, yang cenderung menjadi ketertarikan terbesar bagi penutur asli. Pengembangan ini meliputi aplikasi yang memungkinkan seseorang untuk mencari kata dalam kamus melalui telepon pintar. Selain itu, pengembangan ini juga meliputi aplikasi yang memungkinkan pengguna untuk mengusulkan entri baru atau untuk mengoreksi entri yang sudah ada.

Laman 6 *Apps For Endangered Languages - K International* (<https://www.k-international.com/blog/apps-for-endangered-languages/>) mengilustrasikan pengembangan tersebut. Salah satu aplikasi pertama jenis ini adalah *Ma! Iwaidja*, untuk bahasa Australia, Iwaidja, (<https://appadvice.com/app/ma-iwaidja-dictionary/674860973>). Aplikasi ini meliputi kamus, buku frasa, dan satu komponen yang disebut pembuat kata (*wordmaker*) yang dapat dipakai oleh pengguna untuk memasangkan frasa dan kalimat sehingga tata bahasa dasar bahasa Iwaidja dapat dieksplorasi secara menyenangkan.

Kebanyakan aplikasi mobil yang ada saat ini diciptakan untuk bahasa India Amerika di Kanada dan AS, seperti yang dapat diamati pada laman *First Voices Keyboard app* (<https://www.firstvoices.com/sections/apps>). Aplikasi ini menghadapi satu masalah umum untuk mendukung bahasa kecil yang belum terdokumentasi dengan baik dan khususnya yang terancam punah, yaitu terbatasnya papan tik (*keyboard*) yang dapat menampilkan semua huruf dari alfabet yang digunakan dalam bahasa tersebut.

Selanjutnya, walaupun tidak dibahas secara khusus di bagian ini, segala sesuatu yang telah didiskusikan sehubungan dengan partisipasi, kolaborasi, dan pemanfaatan peran teknologi lebih berperan penting pada upaya konservasi dan revitalisasi bahasa. Konservasi dan revitalisasi bahasa hanya akan berhasil jika komunitas tutur menginginkannya dan terlibat aktif di dalamnya. Dari banyak segi, memiliki aplikasi mobil untuk mengaitkan diri dengan bahasa dan budaya tradisional dengan cara yang menyenangkan di rumah dan di antara teman menjanjikan hal yang lebih penting dan lebih berhasil dibandingkan dengan beberapa jam materi muatan lokal di kurikulum sekolah dasar. Cara ini tidak untuk menyangkal bahwa muatan lokal juga penting dalam

konservasi bahasa karena muatan lokal membantu untuk meningkatkan status bahasa dan menunjukkan dukungan institusi. Namun, bahasa tradisional berkembang di rumah dan di domain pribadi serta umumnya tidak memiliki fungsi di dalam pendidikan dan komunikasi antarkelompok.

## DAFTAR PUSTAKA

- Berez-Kroeker, Andrea L. and Ryan Henke. (2018). Language Archiving. In: Kenneth L. Rehg & Lyle Campbell (eds), *The oxford handbook of endangered languages*. DOI: 10.1093/oxfordhb/9780190610029.013.18
- Bird, Steven and Gary Simons. (2003). Seven dimensions of portability for language documentation and description. *Language* 79:557-582.
- Bird, Steven. (2018). Designing Mobile Applications for Endangered Languages. In: Kenneth L. Rehg & Lyle Campbell (eds), *The Oxford Handbook of Endangered languages*. DOI: 10.1093/oxfordhb/9780190610029.013.40.
- Himmelman, Nikolaus P. (2006). Language documentation: What is it and what is it good for? In: J. Gippert, N. P. Himmelmann & U. Mosel (eds), *Essentials of language documentation*. Berlin: Mouton de Gruyter, 1-30.
- Theimer, Kate. (2011). Exploring the Participatory Archives: What, Who, Where, and Why. Paper presented at the Annual Meeting of the Society of American Archivists, Chicago, Illinois, August 21–26.  
[www.slideshare.net/ktheimer/theimer-participatory-archivessaa-2011](http://www.slideshare.net/ktheimer/theimer-participatory-archivessaa-2011). Accessed 21 August 2018.

# PEMANFAATAN TEKNOLOGI DAN PENINGKATAN KERJA SAMA DALAM DOKUMENTASI BAHASA DAN SASTRA DAERAH

*On the role of technology and team work in language documentation*

**Nikolaus P. Himmelmann**

Universität zu Köln

Pos-el: sprachwissenschaft@uni-koeln.de

## **Abstract**

*Two important features of language documentation, as opposed to more traditional approaches to language description, are its explicit concern for technologies that help to make work on underdocumented languages more efficient and its emphasis on the need for teams of people with different skills and kinds of expertise to work together in documentation and conservation efforts. As shown in the current contribution, these two concerns are closely interlinked and mutually reinforce each other. Online dynamic language repositories ('archives') make it possible for different user groups to access and work with the materials compiled in a language documentation. At the same time, they make it possible for such users to enhance the documentation by adding new data or further annotation and analysis. Similarly, for good documentations it is essential that native speakers and the speech community at large become active partners in the documentation enterprise. This holds even more so for revitalization efforts. Here smart-phone based technology may help to widen the participatory possibilities for native speakers and other interested parties.*

**Keywords:** language archives, dynamic repository, smart phone applications, team work

## **PENDAHULUAN**

Following Himmelmann (2006:1) the goal of language documentation can be defined as creating “a lasting, multipurpose record of a language”, where ‘language’ is to be understood in the broadest possible sense, including all communicative repertoires available in a speech community at a given point in time. The core of a language documentation consists of specimens of linguistic behaviour (ideally video recordings of linguistic interactions) and specimens of metalinguistic knowledge (native speakers’ comments and insights on linguistic forms and practices, often based on elicitation or interviews, but also observed in the documentary core activities of transcription and translation).

In order to be easily usable by different interested parties, and especially ones not involved in the actual documentation efforts, language documentations need to be digitally curated. Section 2 explains what is involved in digital curation, focusing

language archives. Smart phone applications provide a new venue to access, and contribute to, language documentations. This topic is briefly taken up in section 3 which is concerned with the collaborative nature of language documentations. The full potential of a language documentation can only be achieved when people with different skills and kinds of expertise work together.

## 2. On digital curation and ‘participatory archives’

In documentary linguistics, it is common to refer to digital repositories for documentary materials as *language archives*, compare for example (see Berez-Kroeker & Henke (2018:354ff) for more examples and further details):

- *The Language Archive* (<https://tla.mpi.nl>), which includes the *DoBeS Archive* (<https://tla.mpi.nl/resource/dobes-archive/>);
- **the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)**, (<http://www.paradisec.org.au/>);
- **the Endangered Languages Archive (ELAR)**, (<https://elar.soas.ac.uk/>);
- the *Archive of the Indigenous Languages of Latin America* (AILLA, <https://ailla.utexas.org/>).

Using *archive* in reference to such repositories has led to some confusion as the notion of an ‘archive’ is associated by many with a closed, static entity. Traditionally, (paper) archives in fact only housed closed collections such as the writings (letters, manuscript drafts, etc.) of an important person or records of an administrative or political body for a specific year (e.g. the debates of the British parliament in 1957). But language documentations, ideally, should be evolving repositories open to include new data and new annotations. As we will see below, the new concept of *participatory archives* would appear to be more fitting for language documentations. Before clarifying this issue further, it will be necessary to briefly indicate what kinds of data are included in such a repository.

As mentioned above, language documentations consist, on the one hand, of specimens of observable linguistic behaviour, i.e. audio-visual recordings with supplementary materials that help to understand and interpret the specimen at hand.<sup>2</sup> On

---

<sup>2</sup> To keep the exposition simple, I assume for the purposes of this discussion that modern language documentations are largely born digital. That is, audio-video recordings as well as most kinds of additional materials are already produced digitally, the major exception being written notes which

the other hand, they consist of specimens of metalinguistic knowledge, which may be recordings as well (e.g. recording of a speaker and a linguist discussing transcription or translation or recording of a speaker explaining plant taxonomies or kinship relations in a contact language) but often simply consist of handwritten notes taken down when discussing with native speakers linguistic forms or practices. Here as well, supplementary materials may be added.

Supplementary materials, or the *apparatus* in more technical parlance, include a transcription and translation of the recorded speech events, further annotations regarding grammatical or cultural aspects of these speech events, possibly some photos documenting the speech event, handwritten notes made at the time of the recording, and whatever else may be of relevance for understanding and interpreting the recording.

While in principle there are various ways of how to organize the data brought together in a language documentation, it has been found that the most useful basic organization unit is the *session*, i.e. the bundle of all the files pertaining to one communicative event as just listed. Session bundles thus typically include files of different types, i.e. media files of different formats (wav, mpeg, jpg), XML files for annotations and pdfs for notes and other written documents that may have played a role in the event.

Communicative events in the intended sense can be short (e.g. a short joke or story) or very long (e.g. a political meeting extending over several days), the size and complexity of the session varying accordingly. They can also be of different interactional types: everyday conversation, task oriented conversation, narration, elicitation, etc.

For each session, there has to be a metadata file that not only specifies which files belong to the session but also provides basic information on when, where, who, what of the event. Where did the event take place and when? Who participated? What was the basic topic and purpose? What equipment was used for the recording? Which documentation (project) does the session belong to?

The collection of all sessions, i.e. the whole language documentation, also needs metadata, spelling out the goals of the documentation project, who participated, which conventions are followed (orthography, glossing, etc.) and other information needed to

---

typically need to be scanned in order to be included in a digital repository. The inclusion of legacy materials such as analog recordings from earlier times and old fotos and fieldnotes pose special problems not discussed here.

interpret the data in the collection. A major part of these collection-level metadata is an introduction to the language and culture of the community being documented. Here, the distinction between documentation and description in the sense of more in-depth analysis easily becomes blurred. The more analytic materials the collection-level apparatus contains (e.g. full grammars rather than a short sketch grammar, full dictionary with fine-grained subentry structure vs. a working database of lexical entries, etc.), the more useful it is for understanding and interpreting the data collected on the session level.

Table 1 (from Himmelmann 2006:14) summarizes the structure of a language documentation as just outlined.

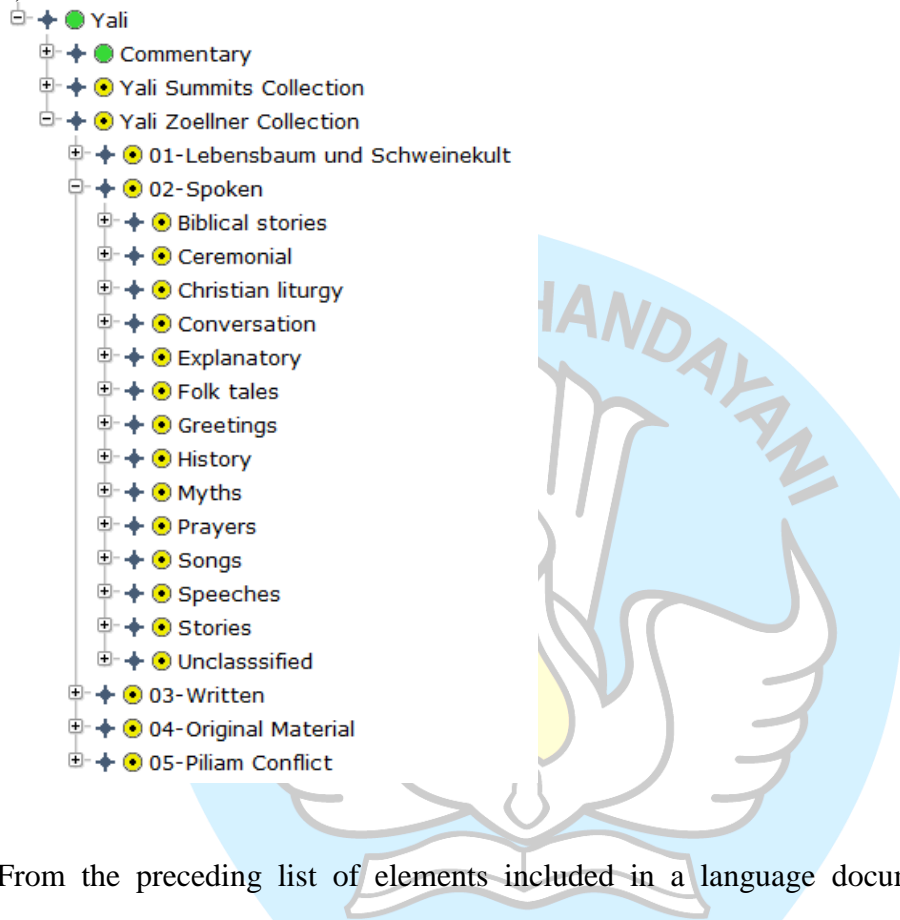
Table 1. *Basic format of a language documentation*

	Per session	For documentation as a whole
recordings/records of observable linguistic behavior and metalinguistic knowledge	Metadata time and location of recording participants recording team recording equipment content descriptors	Metadata location of documented community project team(s) contributing to documentation participants in documentation acknowledgements
	Annotations transcription translation further linguistic and ethnographic glossing and commentary	general access resources introduction orthographical conventions ethnographic sketch sketch grammar glossing conventions indices links to other resources ...

While to the best of my knowledge, all repositories for language documentations to date basically agree on making the session the core of the repository structure, it is an open issue whether a more detailed tree structure is needed that places sessions with a similar content into the same branch. In such a tree, monologue sessions could be distinguished from polylogue ones. Within monologues one could distinguish between

folk stories, personal life stories, procedural texts, etc., as illustrated with the partial tree from the old DoBeS archive in Figure 1. The alternative to such a tree structure is *faceted browsing* which, however, presupposes that the metadata for sessions systematically and reliably contain the information needed for the browsing categories.

**Figure 1:** Example of a tree structure for a language documentation (from old version of DoBeS archive)



From the preceding list of elements included in a language documentation it should be obvious that digital curation is imperative for language documentation. Of course, in principle one could store recording tapes and printed transcriptions, translations etc. together, but to use materials stored in this way would be enormously cumbersome and restricted to people who have physical access to the storage place. The availability of digital video and audio recordings has fundamentally changed the way researchers and other interested parties can access and work with these media types. In fact, they have literally become the anchor point for the prototypical session at the core of a language documentation. That is, transcriptions, translations and other annotations should ideally be time-aligned with the recording not only to make it possible to easily check the original record, but also to be specific and precise as to where exactly a given bit of annotation belongs. With programs such as ELAN (<https://tla.mpi.nl/tools/tla->

[tools/elan/](#)), time-aligned annotation can be performed comfortably and efficiently. These programs have the additional benefit that they store the annotations in the non-proprietary XML format, which is essential for long-term preservation and portability (cp. Bird & Simons (2003) for the importance of *portability* in language documentation).

Digital curation is not only imperative to make a language documentation accessible to other interested parties, but it is also essential for the members of the documentation team who work on annotating and analysing the data. That is, a properly curated digital collection is also a digital research environment where different researchers can access and share data. As media files, especially video, tend to be bulky, language documentations usually comprise several hundred gigabyte or even a terabyte or more, an amount of data not easily handled on standard laptop or desktop computers. Hence, proper digital curation of the data also is in the researchers' very own interest, as otherwise they cannot productively work with the data gathered in language documentation.

Finally, as already indicated above, digital curation is essential for working collaboratively on a given data set. Different researchers working together in a documentation team need to coordinate their annotations and analyses. A prototypical example is the lexical database that will typically be a major constituent of a documentation. Different researchers will add entries when working through transcripts or eliciting items with native speakers. Transcriptions and translations need to be checked and revised. And so on. All of this works best, if the documentation is digitally available.

Importantly, the group of people working on further enriching and enlarging a language documentation need not, and ideally should not, be limited to members of the original documentation team. Instead, other interested researchers and members of the public, especially ones from the speech community being documented, should have the possibility to comment on, add to, and make suggestions for an evolving language documentation by being able to access the documentation and place comments, corrections and suggestions at the specific points they pertain to. An example outside the domain of language documentation would be the possibilities for collaborative writing and publishing provided by *Overleaf* or *Google Docs*. Currently, repositories for language documentations do not yet provide such possibilities for true online



collaboration. But this is clearly the next main step to be achieved in further developing this technology.

Collaboration along the lines sketched in the preceding paragraphs is at the core of the recent proposal to reconceptualize archives as *participatory archives* (see Berez-Kroeker & Henke 2018, section 3, for fuller discussion and references). A simple definition of an participatory archive is provided by Theimer (2011: )

An organization, site or collection in which people other than the archives professionals contribute knowledge or resources resulting in increased understanding about archival materials, usually in an online environment.

Note that this is not an issue specific to language archives, but concerns archiving activities in all domains. For language documentation repositories it would appear that this is the right approach, as it avoids the traditional association of an archive with being static and closed. By their very nature, participatory archives require openness and dynamic update, and hence exactly fit the requirements for a dynamic repository for language documentations.

### **Collaboration and mobile apps**

Continuing the discussion of collaboration at the end of the preceding section, another recent development deserves brief mention, i.e. the use of mobile apps to enlarge the participatory options in language documentations, especially for native speakers.

Good language documentation needs the participation and collaboration of individual native speakers and the community as a whole. “Participation and collaboration” means that members of the speech community participate in the documentation in a way that goes beyond the traditional role of being interviewed and observed, i.e. being the object of research. Instead, good documentary practice requires them to be full active members of the documentation team, proposing their own agenda, wishes and preferences to be taken into account when planning project activities.

In this context, mobile apps provide a way to dissociate participation from physical presence in specific locations. Bird (2018:842) explains:

Speakers of the world’s endangered languages are rapidly gaining access to broadband internet on mobile devices. Meanwhile, social mobile technologies continue to transform the way people work together. I believe that conditions are ripe for the development of a new generation of software for endangered languages. This software

will enable new ways for linguists to collaborate with speakers in ancestral homelands and worldwide diasporas to produce highquality large-scale documentation.

As discussed in Bird’s handbook article (2018), there are many issues involved in developing such apps. Most importantly, they only will be used, if native speakers are involved in their development. Most developments so far, not by accident, are dictionary applications, as these tend to be of greatest interest to native speakers. These include applications that allow one to look up words in a dictionary on a smart phone. But they also include applications that make it possible for users to propose new entries or to correct existing ones.

The website *6 Apps For Endangered Languages - K International* (<https://www.k-international.com/blog/apps-for-endangered-languages/>) illustrates such developments. One of the first apps of this kind was *Ma! Iwaidja*, for the Australian language Iwaidja (<https://appadvice.com/app/ma-iwaidja-dictionary/674860973>). It includes a dictionary and phrasebook and also a component called “wordmaker” with which users can try to put together phrases and sentences and thus playfully explore the basics of Iwaidja grammar.

Most mobile apps to date have been created for American Indian languages in Canada and the USA, as can be gleaned from the website *First Voices Keyboard app* (<https://www.firstvoices.com/sections/apps>). This app addresses another common problem in supporting smaller, underdocumented and especially endangered languages, i.e. the lack of keyboards that represent all letters of the alphabet used for the language in question.

While not specifically the topic of this chapter, it should be obvious that everything that has been said above about participation, collaboration and the enabling role of technology in this regard, holds even more strongly in the case of language conservation and revitalization. Language conservation and revitalization only succeeds when the community wants it and is actively involved in it. In many ways, having mobile apps for playfully engaging with traditional language and culture at home and among friends promises to be much more important and successful than a few hours of *muatan lokal* in the curriculum of primary schools. This is not to deny that the latter are also important in language conservation, as they may help to raise the status of a language and show institutional support. But traditional languages thrive at home and in

the private sphere and typically do not have a function in education or intergroup communication.

## DAFTAR PUSTAKA

- Berez-Kroeker, Andrea L. and Ryan Henke. (2018). Language Archiving. In: Kenneth L. Rehg & Lyle Campbell (eds), *The Oxford Handbook of Endangered languages*. DOI: 10.1093/oxfordhb/9780190610029.013.18
- Bird, Steven and Gary Simons. (2003). Seven dimensions of portability for language documentation and description. *Language* 79:557-582.
- Bird, Steven. (2018). Designing Mobile Applications for Endangered Languages. In: Kenneth L. Rehg & Lyle Campbell (eds), *The Oxford Handbook of Endangered languages*. DOI: 10.1093/oxfordhb/9780190610029.013.40.
- Himmelman, Nikolaus P. (2006). Language documentation: What is it and what is it good for? In: J. Gippert, N. P. Himmelman & U. Mosel (eds), *Essentials of language documentation*. Berlin: Mouton de Gruyter, 1-30.
- Theimer, Kate. (2011). Exploring the Participatory Archives: What, Who, Where, and Why. Paper presented at the Annual Meeting of the Society of American Archivists, Chicago, Illinois, August 21–26. [www.slideshare.net/ktheimer/theimer-participatory-archivessaa-2011](http://www.slideshare.net/ktheimer/theimer-participatory-archivessaa-2011). Accessed 21 August 2018.

