

Documenting Language Use: Remarks on some theoretical and technical issues for language documenters

KIMOTO, Yukinori
Nagoya University, Japan

This paper reviews the documentary linguistics workshop DocLing2013 from the viewpoint of a participant and discusses future prospects on documentary linguistics.¹ The workshop was valuable in that it provided the participants with several key ideas and a rich amount of information crucial for documenting endangered or lesser-known languages. It had a great influence on my own project for the documentation and description of the Arta language; in particular what I learned from the workshop was of great help when I applied for a grant on language documentation. At least two challenges to be addressed are also noted, one of which is relevant to the use of ELAN. I attempt to solve the problem by outlining a systematic use of ELAN in the documentary linguistic context from tier-structuring to the exporting transcripts to other software.

Keywords: DocLing 2013, grant application, interdisciplinary approach, ELAN

1. Introduction
2. DocLing 2013 from a participant's viewpoint
3. Personal benefits from the documentary linguistics workshop
4. Documentary linguistics as an interdisciplinary enterprise
5. Towards a standardization of ELAN
6. Concluding remarks

1. Introduction

I have been working on Arta, an endangered language, spoken by ten fluent speakers and 30–50 second language speakers in the northern part of Luzon in the Philippines. My on-going project (since 2012) concerns the documentation of the

KIMOTO, Yukinori. 2017. "Documenting language use: Remarks on some theoretical and technical issues for language documenters". *Asian and African Languages and Linguistics* 11. pp.79–94. [Permanent URL: <http://hdl.handle.net/10108/89209>]

¹ I would like to thank DocLing lecturers and staff including Peter Austin, David Nathan, and Anthony Duke, from whom I was able to learn a lot about documentary linguistics during DocLing2013. I am indebted to Daisuke Yokomori, Hiromichi Hosoma, Kohei Kikuchi, Mandana Seyfeddinipur, and Sophie Salfner for sharing their knowledge and skills of ELAN. I also thank two anonymous reviewers for comments on this paper. Remaining errors are of course my responsibility. This paper is partially supported by Endangered Languages Documentation Programme (SG0380 "A Documentation and Description of the Arta language").

language (including the recording and analysis of various kinds of discourse), and a description of the phonological system and morphosyntax, which will be submitted as a Ph.D dissertation. When I began this project, I had no systematic knowledge of linguistic documentation, and I struggled to find better ways to record, analyze, and archive the data by myself. It was at that time that fortunately I was informed that a workshop on language documentation (DocLing 2013) would be held at the ILCAA, the Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies. In fact, the ideas I learned during the workshop had a great influence on my project: it led me to reconsider the ways of documenting the language, from the choice of microphone to future plans for archiving the data.

In this paper, I will review DocLing 2013 from the viewpoint of a participant, and will discuss prospects for language documentation research. In §2, the Documentary Linguistics Workshop 2013 will be reviewed focusing on the lectures and some features characteristic of DocLing; in §3, I will remark on the benefits I have had from the workshop, in particular, I focus on how it has had an impact on my research, in particular, when I applied for an international grant on endangered language documentation. In §4 and §5, future prospects for documentary linguistics, concerning both theoretical and practical aspects, will be discussed.

2. DocLing 2013 from a participant's viewpoint

DocLing 2013 was held for six days in 11th–16th February, 2013, at the Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies. Lectures were given by Peter Austin (SOAS, University of London), David Nathan (SOAS, University of London²), Anthony Jukes (CRLD, La Trobe University), Sonja Riesberg (University of Cologne), and Hideo Sawada (ILCAA, Tokyo University of Foreign Studies). Participants devoted almost half of the time to group projects on language documentation, collaborating with the following native speakers whose languages were not familiar to the participants: Tshering Tashi,³ Namgay Thinley,⁴ Zhargal Badagarov,⁵ and Kristian Walianggen.⁶ ILCAA staff also assisted the participants during the workshop. The workshop also encouraged the participants to interact with each other and with staff; participants were given sufficient break times to communicate with each other informally.

Let me briefly review the lectures that I found quite impressive; the program of

² Currently Co-ordinator of the Centre for Australian Languages and Linguistics, at the Batchelor Institute of Indigenous Tertiary Education.

³ Royal Society for Protection of Nature, Board Director, Dzongkha native speaker (Bhutan)

⁴ Dzongkha Development Commission, Senior Research Officer, Dzongkha native speaker (Bhutan)

⁵ Buryat State University, Lecturer, Buryat native speaker (Buryatia, Russian Federation)

⁶ Center of Endangered Language Documentation, State University of Papua, Language Consultant, Yali native speaker (Papua, Indonesia)

DocLing 2013 is given in Figure 1⁷:

	10:00-11:20	11:40-13:00	14:00-15:20	15:40-17:00
Day 1	Introduction to language documentation (PA)	Planning language documentation and group projects (AJ, DN)	Public lecture: "Rethinking language documentation and support for the 21st century" (PA, DN, AJ)	Ethics and working with communities (PA, guest presenter Anna Berge)
Day 2	Audio & video (DN, AJ)	Audio & video (AJ, DN)	Set up groups & projects (DN)	Group projects design & report (group & staff)
Day 3	Fieldwork techniques & elicitation (SR)	Recording with consultants (groups & staff)	Software for documentation: survey (PA, AJ)	Software for documentation: practical demonstration (AJ)
Day 4	Data management (DN)	Data management (DN)	Multimedia & mobilisation (AJ, DN)	Photography for documentation (HS)
Day 5	Archiving (DN)	Group project development (group & staff)	Group project development (group & staff)	Individual clinic / Group project (group & staff)
Day 6	Individual clinic / Group project (group & staff)	Group project finalisation (group & staff)	Project presentation (groups)	Project presentations (groups), Wrap-up (all)

Fig. 1 Program of DocLing 2013

Overview of Documentary linguistics (Day 1, 1st and 3rd classes, Austin 2010a "Current Issues in Language Documentation"): The lectures and the reading material distributed before the workshop enhanced our knowledge of language documentation or documentary linguistics, including the definition of documentary linguistics, differences between documentary and descriptive linguistics, the current situations of documentary projects in the world and the availability of research fundings, such as ELDP (Endangered Languages Documentation Programme) at SOAS, University of London, the Volkswagen Foundation DoBes project (Dokumentation bedrohter Sprachen), and DEL (Documenting Endangered Languages), an interagency program of the National Science Foundation and the National Endowment for the Humanities. Since this academic field has few introductory textbooks, we found his lectures and the paper to be excellent introductions to the field.

Ethics in linguistic documentation (Day 1, 1st and 3rd classes, Austin 2010b): Although linguistic fieldwork presupposes a relationship between researchers and members of the community, the field of linguistics had not begun to address ethical issues seriously until fairly recently. This lecture and the reading material addressed

⁷ From <http://lingdy.aacore.jp/en/activity/docling/2013.html> (accessed at Sep 9th, 2016); PA: Peter Austin, DN: David Nathan, AJ: Anthony Jukes, SR: Sonja Riesberg, HS: Hideo Sawada.

such ethical issues as legal aspects (copyright, intellectual rights, and so forth), principles of informed consents, and how to establish a good relationship with a community. In Japan, grant applications often require an explicit statement about ethical matters, but we do not have opportunities to focus on such matters in detail in academic institutions. It was thus fortunate that we were able to discuss the matter in this class and with other participants.

Audio and video recording, photography (Day 2, 1st and 2nd classes, Day 4, 4th class, Nathan 2010): Probably one of the most significant academic impact of the field of documentary linguistics in general is that it has made it possible to discuss the technical and practical issues on audio and video recording as an academic pursuit, and has made it possible to accumulate and share the knowledge. In this lecture and the reading material, we learned about epistemological principles of recording (especially the importance of human ears in evaluating recordings), the basics of audio and video recordings, and the typology of microphones, we also saw practical demonstrations of using microphones and learned about various ways of taming noises. All of the participants realized how important it is to choose appropriate microphones based on the conditions of the field, and to adjust the settings carefully for better recordings.

Software for documentation (Day 3, 3rd and 4th classes): Two lectures were given concerning software used for language documentation. The first lecture provided a brief overview of various pieces of software, including the purposes of using each one, and their advantages and disadvantages. The second lecture, the participants practiced using ELAN (annotation software for media files) and Toolbox (software for conducting grammatical analysis and building a lexicon). It was of great help for those who wanted to use software in language documentation. However, the lecture was only the first step for using the software; thus slightly more systematic teaching would also be necessary.

Data management (Day 4, 1st and 2nd classes): On Day 4, two lectures were given on data management, in which participants were systematically taught how to manage various kinds of data files (such as .eaf (ELAN), .mp4, .wav, and so on). One of the important ideas relates to the labeling of file names. File names are identifiers that should uniquely distinguish one file from another. It is thus advantageous to give a simple file name with ASCII characters, which are structured systematically across files (e.g. arta0031). Only a limited amount of semantic information should be included in the file names; meaningful information should mainly be stored in a metadata file, which is linked to the files using an identifier. This idea has influenced my project in terms of my management of various kinds of files. I still remember “Always backup the data!” and “A filename is an identifier, not a data container”.

In sum, the workshop provided us with a rich amount of important ideas for language documentation, most of which were not taught in universities, probably for the simple

reason that they have not been established as a discipline.

3. Personal benefits from the documentary linguistics workshop

I received many benefits from the documentary linguistics workshop, the first of which was the knowledge of how to produce strong grant applications. I applied for an Endangered Languages Documentation Programme (ELDP) grant funded by SOAS, University of London, and was awarded a one-year grant titled “A Documentation and Description of the Arta Language” (SG0380; Yukinori Kimoto, 2015–2016). The grants require the applicants to write a long description of the project, and the ELDP expects them to be familiar with documentation methods. Figure 2 shows the actual questionnaires in the 2015 version of the ELDP grant. For this application form, specific descriptions are expected for several questions, including the summary of the project (Q10, max. 100 words), the description of specific outcomes (Q11, max. 750 words), the detailed description of the project including language context, documentation methods, community context, and research outcomes (Q12, max. 2000 words, thus around 400 words per subcategory), ethical aspects of the project (Q13, max. 750 words), the project work plan (Q14, max. 750 words), and financial details of the support requested (Q15). One can easily imagine that applicants are required to have a systematic knowledge of language documentation. Thanks to the documentary linguistics workshop, I understood what the applicants were expected to describe. I reviewed the slides and papers that had been shared during the documentary linguistics workshop, and was able to compose my proposal in a clear and suitable manner.

<p>Q1 Applicant details (name, address, his/her current position, etc.)</p> <p>Q2 Host institution which will administer the award (including a letter of acceptance by head/dean of the institution)</p> <p>Q3 Title of the project</p> <p>Q4 Duration of the project</p> <p>Q5 Proposed start date</p> <p>Q6 Previous ELDP grants? (whether s/he have held a grant from ELDP)</p> <p>Q7 Is this re-submission?</p> <p>Q8 Other funding applications</p> <p>Q9 Commitment to the project (the intended average number of hours per week, etc.)</p> <p>Q10 Summary of the proposed project (Maximum 100 words)</p> <p>Q11 Specific outcomes (Maximum 750 words)</p> <p>Q12 Detailed description of the project (Maximum of 2,000 words across the 4 headings) (i) Language context, (ii) Documentation methods, (iii) Community context, (iv) Research outcomes</p> <p>Q13 Ethical aspects of the project (Maximum 750 words)</p> <p>Q14 Project work plan (Maximum 750 words)</p> <p>Q15 Financial details of the support requested (Specify the details with calculations, and state the justifications of the items.) (a) Replacement teaching costs/research salaries, (b) Technical staff, (c) Language consultants, (d) Equipment, (e) Travel and subsistence, (f) Consumables, (g) Other costs</p> <p>Q16 Referee (Referee details)</p> <p>Q17 Curriculum Vitae</p>
--

Fig. 2 ELDP grant application form (c.a. 24 pages)

Second, and perhaps more importantly, the workshop played a significant role in connecting field researchers in Japan and across countries. The one-week workshop includes a group session as an integral part of it, in which the participants were encouraged to conduct collaborative work. As mentioned above, during the longer breaks between lectures we communicated with each other informally. Our social relationships established in DocLing 2013 is still maintained.

Furthermore, since the workshop was strongly connected to SOAS, it was easy for me to extend my social network to people in SOAS. I attended two workshops held at SOAS: “Plants, Animals, Words”: a workshop on ethnobiological, linguistic, and anthropological studies (September, 2013), and an ELDP training workshop (September, 2015), and I shared fruitful experiences with the other participants, particularly with young field researchers.

In the next two sections, the scope of discussion is broadened into documentary linguistics as a whole. Focusing on the documentation of language use, I will explore the theoretical significance and a practical challenge of documenting language use. In §4, I reconsider the interdisciplinary nature of documentary linguistics, and will suggest that any documentary project without theoretical considerations would be useless to researchers in related fields; we should learn related fields wherein the documentation of language use is of great significance. In §5, a practical issue of annotating discourse data with ELAN, annotation software, is addressed. Based on the technical workshop I gave in Kyoto University in Japan, some basic method and advanced use of ELAN will be explicated.

4. Documentary linguistics as an interdisciplinary enterprise

Although documentary linguistics is supposed to “work in interdisciplinary terms” (Austin 2010a: 13), it seems that there are so far few interdisciplinary projects involving or applying documentary linguistics, with the primary focus of most other research in the field remaining on the interactions between language documentation and language description (or documentary vs. descriptive linguistics).⁷ If language documentation is aimed at the preservation of primary data for long-term storage in interdisciplinary terms, trainings on language documentation should show participants how our data can contribute to research in other disciplines. Whether or not the documenters have such knowledge may seriously affect the quality of data collected.

Consider a hypothetical situation in which a linguist used an IC recorder to record a procedural text of traditional cooking, and archived the data with the transcription: “*we put that here, and do it like this, or like this, and after that, we put it there*”. It would be no problem if he just wants to write a grammar, but language documenters should note

⁷ In fact, in DocLing 2013, all the lecturers and staff were linguists, and almost all the participants in DocLing 2013 were linguists. We participants were thus quite unclear as to how we can develop our work “in interdisciplinary terms” possibly with researchers of other fields.

that non-verbal information might be a crucial key to understand their cultural practices in cooking when none of the community member lost such cultural knowledge.

It is necessary for language documenters to at least be familiar with other (broad) areas of studies in languages because ignorance of other academic fields results in data collections that are totally useless to researchers involved in linguistic or come in other fields. There is no theory-neutral documentation (**theory-ladenness of observation**). After all, we are the one who “constructs” and “frames” the data from a particular viewpoint.

Actually, the primary data of language use are not useful only for descriptive linguistics, but are also of great use for other fields. Some typical fields to which the primary data of language use can contribute are shown below:

- usage-based approach in cognitive linguistics (Langacker 1987, Bybee 2010)
- gesture studies in psychology (Kendon 1988, McNeill 1992, Clark 1996, Kita 2003)
- research on language socialization in anthropology (Ochs et al. 2001, Takada 2012)
- conversation analytic studies in sociology (Sacks, et al. 1974, Schegloff 2007)
- interactional linguistics—an interdisciplinary enterprise between conversation analysis and functional linguistics (Couper-Kuhlen and Selting 2001, Hayashi 2003)

The attempt to apply documentary linguistic studies to these fields will reveal illuminating facts on language, cognition, social interaction and culture.⁸ If practitioners of language documentation are informed of such orientations, their data are potentially of great use to these fields.

5. Towards a standardization of ELAN

The second challenge of documenting language use is relevant to a technical issue, concerning the standardization of ELAN.⁹ It seems that the practical skills in ELAN that are necessary for language documentation are not always taught in the context of language documentation trainings. Since DocLing ended, I learned more about how to use the software through trial and error with the help of the manual and other users of ELAN; this has sometimes required me to radically change the format of some of my annotated files to make them consistent.

⁸ Mark Dingemanse, Nick Enfield, and their colleagues, for example, surveyed typological variations of so-called “other-initiated repair” (Dingemanse and Enfield 2015), revealing remarkable commonalities across languages/cultures.

⁹ ELAN was developed by the Max Planck Institute for Psycholinguistics; it was designed for a number of different users such as linguists, language documenters, gesture and sign language researchers, and for their collaborations (Brugman and Russel 2004). It is powerful and enables various methods of annotation, so that it has become an essential tool for language documentation at the moment.

In this section, I attempt to provide one model for how to teach ELAN to language documenters and the students of linguistics in general. The model tutorial presented below was gradually modified through the process of writing an ELAN tutorial for language documentation (Kimoto (to appear)), and technical workshops on ELAN, which was held three times in the summer of 2016 designed for linguistics students at Kyoto University, Japan.

For ELAN users in the context of language documentation or linguistic analysis, the following skills and types of knowledge are especially relevant:

1. A basic method of annotating media files.
2. The conceptual difference between linguistic types and tiers, and their practical usage.
3. Advanced usage: automatic numbering, word-breaking, controlled vocabulary

Two-hour workshop was designed to include the above contents: the outline of the workshop is given in Figure 3. Prior to the workshop, participants had finished installing ELAN, and had been provided with a movie file (.mp4), and a sound file (.wav) extracted from the movie file. The movie file had been recorded in my field session on the Arta language.

In the technical workshop, I demonstrated the method to use ELAN in my field situation, where a researcher is annotating speech in Arta, spoken by Speaker A, and translating it into Ilokano (a lingua franca widely used in the northern Luzon area of the Philippines) and English. Participants practiced ELAN within the exemplar situation.

1. Basics of ELAN:
 - 1.1. A simple method of annotation
 - 1.2. Using ELAN transcriptions as a corpus: searching across annotation files
2. Constructing tier structure
 - 2.1. Conceptual difference between linguistic types and tiers
 - 2.2. Setting up linguistic types
 - 2.3. Adding tiers
 - 2.4. Saving as a template
 - 2.5. Adding participants from a single participant template
3. Advanced uses of ELAN
 - 3.1. Synchronizing video and audio data
 - 3.2. Notes on video data
 - 3.3. Autonumbering of annotations
 - 3.4. Controlled vocabulary
4. Exporting
 - 4.1. Exporting as an interlinearized text file
 - 4.2. Exporting to FLEx (FieldWorks Language Explorer)

Fig. 3 Technical workshop on ELAN for linguistic analysis

In what follows, I provide one model for teaching ELAN to language documenters, which was modified based on the workshop I gave to linguistics students.

Session 1: Basics of ELAN

A simple method of annotation As a first step, participants are directed to notice how simply they can annotate media files. After the movie and sound files are imported into ELAN, the participants are directed to annotate files in the simplest way (without making any change in linguistic type and tier). They learn the following points:

- how to use shortcut keys. In particular, annotations cannot be moved or expanded/contracted without using the ALT key (Table 1).
- media player options; e.g. arranging the playback speed
- representation of the waveform (it is often the case that if the sound is recorded in a low volume, the waveform is not visible clearly): right-click on the waveform viewer > vertical zoom > select 300% (Note that waveforms are not shown if the media type is video!)

Table 1 Shortcut keys in ELAN

Start/Pause the playback:	CONTROL + SPACE
Play the selected interval:	SHIFT + SPACE
Enter an annotation into a selection:	ALT + N (active tier only)
Edit a selected annotation:	ALT + M (active tier only)
Save the current annotation:	ENTER
Move an annotation:	ALT + drag the middle of the annotation
Expand/contract the length of an annotation:	ALT + drag the border of the annotation

Using ELAN transcriptions as a corpus: searching across annotation files The next session introduces methods of searching in ELAN, particularly because the ability to search across files is one of the main advantages of using ELAN; this makes it possible to use our data as a corpus. Here just a short introduction may be provided, including a method of defining the corpus (“define domain”), and the availability of regular expressions (with reference to Mosel 2015).

Session 2: Constructing tier structure

When using ELAN, it is essential to understand the conceptual difference between linguistic types and tiers. But this is not always easy for learners to understand quickly. In the workshop I gave, I used illustrations to make it easier to grasp the ideas (Figure 4).

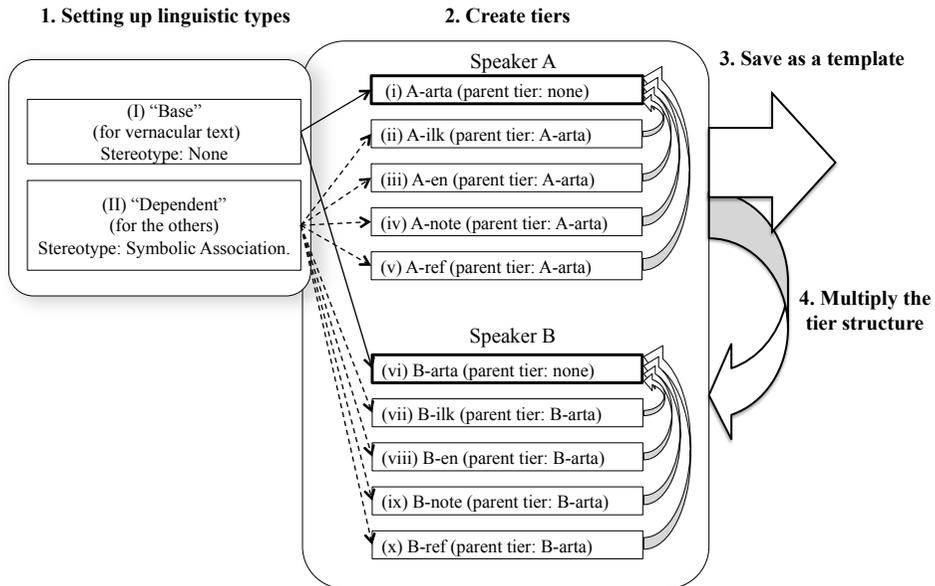


Fig. 4 Tier-structuring process

In the workshop, vernacular texts (in Arta) were assigned to the “base” linguistic type, and the others (Ilokano translation, English translation, grammatical/lexical notes, reference number) the “dependent” linguistic type. Participants were directed to create a tier structure for one speaker, and to save it as a template for further use of the same tier structure. They also learned to multiply the tier structure in case more than one speaker is relevant to the transcription. This process seems more advantageous than manual additions particularly because users can create the same tier structure in quite a simple way.

Conceptual difference between linguistic type and tier I illustrated the conceptual difference between linguistic type and tier by explaining the following to students. Suppose you need five kinds of information for each annotation, that is, (i) a transcription of the object language (here “Arta”), (ii) a translation in the lingua franca (Ilokano), (iii) an English translation, (iv) grammatical/lexical notes, and (v) the utterance ID. In this case, these five layers should be prepared in the setting. But on a more abstract level, they can be classified into two general types: (I) one in which you can freely change the intervals and positions within the timeline (that is, (i) Arta), and (II) another in which the intervals and positions are dependent on the first tier type (that is, (ii)–(v)). These two general tier types are called **linguistic types**, and the five actual layers are called **tiers** in ELAN. If the media file to annotate includes three speakers, 15 “tiers” should be prepared, whereas two “linguistic types” are sufficient for the purpose.

Setting up linguistic types and adding tiers. After reaching a conceptual understanding of the two different levels of tier structure, participants learn other key concepts such as “stereotype”, “parent tier”, “symbolic association”, and inheritances between linguistic types and tiers, guided by the illustration in Figure 4.

Saving as a template. After a tier structure for one speaker is constructed, the tier structure is saved as a template (File > Save as Template). Doing so eliminates the need to re-create the same tier structure for each annotation file, and allows for consistent tier structures across files.¹⁰

Adding a new participant. If the media file being annotated has two or more speakers, the tier structure for one speaker can be multiplied in a relatively simple manner (rather than manually adding tiers). Click on Tier menu > Add New Participant... (then a dialog window will appear). In the dialog window, (i) check if the tier structure to copy is selected; (ii) enter the name of the speaker to add in “Specify the new participant”; (iii) select the relative position of the name value specified in the label in “Specify whether the suffix or prefix of the selected tier structure(s) to be changed” (for a new tier structure like “B-arta, B-ilk, ...” from “A-arta, A-ilk, ...”, select “prefix”, or for “arta-B, ilk-B, ...”, select “suffix”); (iv) enter the name value of the speaker in the original tier structure (e.g. A) in “Enter the value to be replaced”; (v) enter the name value of the new speaker (e.g. B) in “Enter the new value for replacement”; (vi) select “OK”, and the result will be shown in the timeline viewer.

Session 3: Advanced uses of ELAN

During the workshop I organized, participants worked with some of the more advanced uses of ELAN, such as annotation autonumbering, controlled vocabulary, and word/morpheme-breaking; they also practiced synchronizing multiple media files. These advanced functions are not necessarily required when using ELAN, but are useful in some cases. Here I just make some short remarks on the synchronization of media files.

Synchronizing video and audio data. Although ELAN provides ways to synchronize more than one media file (e.g. video-video, video-audio), researchers should first consider whether this is really necessary. If a single session has been recorded with two videocameras, it would make sense to synchronize those two video files. However, if one hopes to synchronize video and audio files (e.g. suppose a situation in which session was both video-recorded with a camcorder, and audio-recorded with an IC-recorder), one should reconsider whether that is the only option available. Since synchronization in ELAN can be a troublesome task, I always

¹⁰ Even if tier structures are inconsistent across files, ELAN does work. The problem arises when you search across files, however. One of the advantages of cross-file searching is that a particular tier or speaker can be specified. If tier structure are inconsistent across files, you cannot search across files according to these attributes.

record sessions with a videocamera and an external microphone; this seems sufficient to get a clear sound with less noise.

Various pieces of software can be used to extract .wav files from .mp4 or other video files (e.g. Avidemux (Windows), QuickTime Player 7 (Mac), Audacity (Windows, Mac)); using .wav files enables ELAN to show waveforms along the timeline.¹¹

Session 4: Exporting interlinearized texts into FLEx (FieldWorks Language Explorer)

FLEx or FieldWorks Language Explorer is software used for building a lexicon, analyzing texts, and studying morphology and syntax, distributed by SIL International. Although it is also powerful software, and ELAN and FLEx have an interface for exporting/importing each other's files, it is quite a complex procedure. The manual "Working with ELAN and FLEx together" is available under "Third party resources" on the ELAN website. It begins with instructions to set up a tier structure and to annotate a media file in a special way, which do not really answer what we want to know; if you already have several annotated ELAN files to export into FLEx, this method does not work at all.

This is the background motivation to explicate what is probably the most realistic way to export annotated ELAN files into FLEx below in detail.

Step 1. Add the linguistic type of the title. FLEx cannot import the data properly if the labels of linguistic types do not follow a designated format and if the data do not have a title tier. Here you need to add a linguistic type for the title of the annotated file. Go to "Type" > "Add Linguistic Type" and create a new linguistic type with the name "Title", and the stereotype "None". Next, go to "Tier" > "Add tier" and create an additional tier with the name "Interlinear-title-[ISO639]" (see also Table 2).

Step 2. Assign participant information to the annotations. Interlinear texts in FLEx are not adopted to multi-participant discourses so far. If the session annotated is a conversation by more than one speaker, such information on the speakers should be put in the "notes" tier. Create a new tier with the dependent linguistic type, and add new annotations of speaker information below each utterance (e.g. "A", "B", etc. for each annotation). It is easy to make empty annotations automatically by clicking "Tier" > "Create Annotations on Dependent Tiers..." and entering the speaker information, "A" or "B", over the same tier by clicking "Tier" > "Label And Number Annotations..."

Step 3. Change the tier names. FLEx requires tier names to follow a particular format to be able to read them properly. Change the names of relevant tiers (Table 2). In particular, ISO-639 codes for an object language or vernacular language (e.g. atz for Arta), and a translation language (e.g. en for English) are required.¹² You can consult

¹¹ Note that the waveform does not appear if the annotation file is linked only to video file(s).

¹² It seems that this information should correspond to the ISO codes you used when a FLEx project was built.

*Ethnologue*¹³ or even *Wikipedia* to get the ISO codes.

Table 2 Tier names for FLEx exporting

	Format	Example
Title	Interlinear-title-[ISO639-of-object-lang.]	Interlinear-title-atz
Transcription	[speaker]-txt-[ISO639-of-object-lang.]	A-txt-atz
Translation	[speaker]-gls-[ISO639-of-translation-lang.]	A-gls-en
Speaker	[speaker]-note-[ISO639-of-translation-lang.]	A-note-en

The whole tier structure modified here is shown in Figure 5, where the difference between the one constructed here and the one illustrated in Figure 4 is shown in bold.

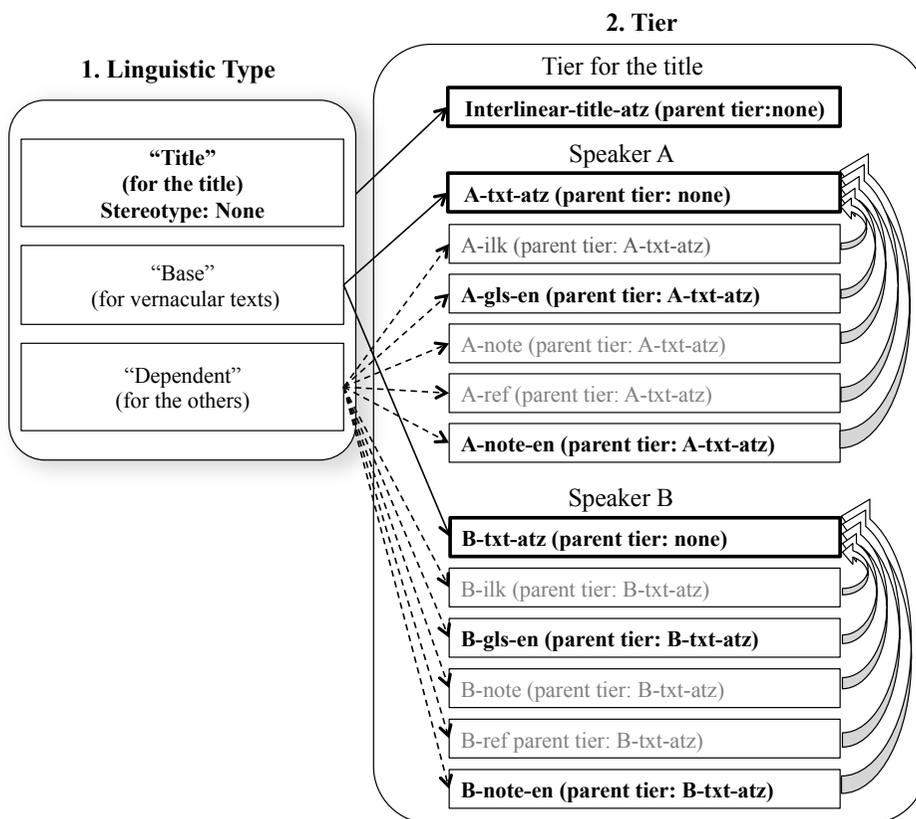


Fig. 5 Tier structure for FLEx exporting

¹³ Online version: <https://www.ethnologue.com/>.

Note that if the transcription tier is not a parent tier, but if another tier (e.g. a reference number tier) is a parent tier within the tier structure, then the parent-child relation must be reversed. ELAN does not seem to have the function of changing the tier structure directly, but the “copy tier” function can be used to do this. First, copy the transcription tier (i.e. a dependent tier), selecting “Transcription (no parent)” to make it an independent tier, then copy the original parent tier (e.g. the reference number tier), placing it under the new parent tier. After editing the tier structure in this way, delete the original tier structure.

Step 4. Annotate the title. Create one annotation of the `Interlinear-title-xxx` tier in an arbitrary location, and enter the title of the file, which may correspond to the filename (e.g. “arta0032”), or which may be a more informative title (e.g. “building a traditional house”). This annotation becomes the title of the interlinear text in FLEEx (but sometimes this does not work, and you might need to input the title in FLEEx manually!).

Step 5 Export Choose File > Export as > FLEEx file and an “Export as FLEEx File” window will appear:

Screen 1 Setting up the Title/Base linguistic type and tier(s)

- In the upper box “Element mapping”, select “Export interlinear text tier” (but not “Export paragraph tier”), and then select “Interlinear-title-xxx” in the drop-down menu. Confirm that “phrase” in the “FLEEx Element Name” column corresponds to “Base” in “Corresponding Tier Type” column. If not, select “Base” from among the candidates.
- In the lower box “Select tiers to be exported”, select all of the transcription tiers (e.g. “A-txt-atz” and “B-txt-atz” in Figure 5) in the “phrase” column, and unselect the other tiers.

Screen 2 Setting up the Dependent linguistic type(s) and tier(s)

- In the upper box, check if “phrase_item” in the “Item Mapping” includes the “Dependent” type.
- In the lower box (“Select tiers to be exported”), select all of the dependent tiers necessary (e.g. translation tiers: “A-gls-en”, “B-gls-en”; and participant-value tiers: “A-note-en”, “B-note-en”), and unselect the others.

Screen 3 Nothing to arrange. Click “Next”.

Screen 4 Specify the location for saving the file where FLEEx can access it. Click “Finish”.

Now the file has been exported. To import the file in FLEEx, open the project where the data will be imported. Select “View” > “Texts & Words” > “Interlinear Text” (or select “Texts & Words” on the lower left side of the window). Go to “File” > “Import” > “FLEEx Text Interlinear”. Specify the location of the file and open it.¹⁴ Now you can

¹⁴ At present (Sep.11th, 2016), there is a bug such that the text cannot be recognized as a vernacular language. In

view data from ELAN in FLEx.

6. Concluding remarks

In this paper, I reviewed the documentary linguistics workshop DocLing 2013, where I looked back at the classes that were particularly beneficial from a participant's perspective. I also described how the workshop had an influence on my research; the lectures on documentary linguistics that I attended as part of the workshop were of great help to me when I applied for the Endangered Languages Documentation Programme (ELDP) grant. Finally, I mentioned some current challenges in documentary linguistics, focusing on the documentation of language use, i.e. the lack of interdisciplinary approach, and the absence of standardization of ELAN. For the latter problem, I attempted to explicate a better way to use ELAN with a special focus on tier-structuring and the interface between ELAN and FLEx. I hope the discussions provided here will provide some perspectives and insights for the further development of documentary linguistics in general and projects around the world.

References

- Austin, Peter K. 2010a. "Current issues in language documentation". In Peter K. Austin (ed.) *Language Documentation and Description 7*. London: SOAS. pp.12–33.
- Austin, Peter K. 2010b. "Communities, ethics and rights in language documentation". In Peter K. Austin (ed.) *Language Documentation and Description 7*. London: SOAS. pp.34–54.
- Brugman, Hennie, and Albert Russel. 2004. "Annotating multi-media / multi-modal resources with ELAN". In *In: Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Clark, Herbert. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Couper-Kuhlen, Elizabeth and Margret Selting. 2001. "Introducing interactional linguistics". In Margret Selting and Elizabeth Couper-Kuhlen (ed.) *Studies in Interactional Linguistics*. London: SOAS. pp.1–22.
- Nathan, David. 2010. "Sound and unsound practices in documentary linguistics: Towards an epistemology for audio". In Peter K. Austin (ed.) *Language Documentation and Description 7*. London: SOAS. pp.262–284.
- Dingemanse, Mark and Nick J. Enfield. 2015. "Other-initiated repair across languages: Towards a typology of conversational structures". *Open Linguistics* 1(1). pp.96–118.
- Hayashi, Makoto. 2003. *Joint Utterance Construction in Japanese Conversation*. Philadelphia: John Benjamins Publishing.
- Kendon, Adam. 1988. "How gestures can become like words". *Cross-Cultural Perspectives in Nonverbal Communication* 1. pp.131–141.
- Kimoto, Yukinori. to appear. "Gengo dokyumenteesyon notameno ELAN siyoo (ELAN for language documentation)". In Hosoma, Hiromichi and Kohei Kikuchi (eds.) *ELAN Nyuumon (An Introduction to ELAN)*. Tokyo: Hituzi Syobo Publishing.
- Kita, Sotaro (ed.) 2003. *Pointing: Where Language, Culture, and Cognition Meet*. New Jersey: Lawrence Erlbaum Associates.

such a case, select all of the text and choose the language value in the drop-down menu of the languages appearing in the second-from-the-top row.

- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Vol.1 Theoretical Prerequisites*. Stanford: Stanford University Press.
- McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- Mosel, Ulrike. 2015. "Searches with regular expressions in ELAN corpora". [Online at <https://tla.mpi.nl/tools/tla-tools/elan/thirdparty/> under the description of "Regular expressions for ELAN users", accessed on 2016-09-12].
- Ochs, Elinor and Bambi Schieffelin 2001. "Language acquisition and socialization: Three developmental stories and their implications". In Alessandro Duranti, Elinor Ochs and Bambi B. Schieffelin (eds.) *Linguistic Anthropology: A reader*. Chichester: Blackwell. pp.263–301.
- Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson 1974. "A simplest systematics for the organization of turn-taking for conversation". *Language* 50.4. pp.696–735.
- Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction: Vol.1: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
- Takada, Akira. 2012. "Preverbal infant–caregiver interaction". In Alessandro Duranti, Elinor Ochs and Bambi B. Schieffelin (eds.) *The Handbook of Language Socialization*. Chichester: Blackwell. pp.56–80.