You’ve made a recording. Now what?

- You probably need to transcribe it.
- You may need to translate it.
- You may want to add other information.

Some tools will help you transcribe. ELAN and Transcriber are two that linguists are using these days.
ELAN

• “ELAN (EUDICO Linguistic Annotator) is an annotation tool that allows you to create, edit, visualize and search annotations for video and audio data.”
• links text annotations with audio and/or video data.
• one audio stream, up to four video streams
• ELAN files can be exported in a variety of formats (including to Shoebox/Toolbox for interlinearisation, then reimported)

What can’t ELAN do?

• It can’t do your transcription
• It can’t do your analysis
• It can’t keep you organised
• It can’t (by itself) make a viewer for community members
• It isn’t (unfortunately) very easy to learn
What can ELAN do?

- It can help with transcription and translation
- It can help with your analysis by presenting your data
- It can help keep you organised by linking the media and data files together
- It can help you find things in your data
- It can help if making a product for community members (text, subtitled video)
Tiers

- Tiers are where you put your annotations
- Tiers can contain many kinds of annotations, some of the most obvious are:
  - IPA transcription
  - practical orthographic transcription
  - free translations into languages of wider communication
  - morphemes and gloss
  - gesture annotation
  - grammar notes
  - any other information which seems relevant

ELAN – pros and cons

- Handles most audio and video formats
- Powerful for annotating and searching
- Good compatibility with Toolbox
- Good exports for web video etc via CUPED or other tools, YouTube
- Prospects for development
- Multi-platform, open-source
- Difficult to get started – steep learning curve
- No inbuilt tools for interlinearising or lexicon building
- *Too* powerful/flexible – temptation to add zillions of tiers, gets cluttered and confusing
Transcriber

- Transcriber is a tool for assisting the manual annotation of speech signals.
- It provides a user interface for segmenting long duration speech recordings, transcribing them, and labeling speech turns, topic changes and acoustic conditions.
Transcriber pros and cons

- Relatively easy to set up and use
- XML format for easy file exchange
- Handles most audio formats
- Multi-platform, open source

- Lacks video support
- Overlapping speech tricky to handle when exporting to Toolbox
- Not (really) designed for linguists – unlikely to integrate with linguistic analysis tools in the future

You’ve transcribed. Now what?

- Grammar analysis
- Lexicon building
- Cultural/ethnographic notes
- ???

Tools that help you do some of these things:
  - Toolbox
  - Fieldworks Language Explorer (FLEEx)
  - both from SIL
Toolbox

- Toolbox is a data management and analysis tool for field linguists.
- It is especially useful for maintaining lexical data, and for parsing and interlinearizing text, but it can be used to manage virtually any kind of data.
Toolbox pros and cons

- Tried and tested
- (Relatively) easy to use after some initial study
- Large and helpful user community
- Interoperability with ELAN
- Can produce printed or online dictionaries with MDF or Lexique Pro
- Standard Format (backslash codes) not really well-structured
- ‘End of life’? It is very old, not being developed actively
- Limited interaction with media files
- Mac only under emulation

Fieldworks Language Explorer

- “FieldWorks is a set of software tools that help manage cultural and linguistic data from initial collection through submission for publication”
- It can be used to record lexical information and develop dictionaries.
- It can interlinearize text.
- The morphological parser provides the user with a way to check the grammatical rules they have recorded against real language data.
- The grammar information can also be compiled in an automatically generated grammar sketch.
This is a simple sentence about seeing a green mat.

I see my green mat.
FLEEx pros and cons

- Better data structure than Toolbox - XML
- Very powerful parsing and grammatical analysis tools
- Designed to hold all your linguistic and cultural data and notes
- Poor handling of media
- Large application, memory hog
- Windows only
- Poor integration with Toolbox

Another dictionary tool – We Say

- WeSay helps non-linguists build a dictionary in their own language.
- It has various ways to help native speakers to think of words in their language and enter some basic data about them (no backslash codes, just forms to fill in).
- Designed for teamwork – one ‘advanced’ user does the complicated set-up work, very simple interface for other users
We Say pros and cons

- Very simple to use
- Will run on netbooks and other low-powered machines
- Good data structure
- Easy export via Lexique Pro for print/web
- No tools for interlinearising or analysis
- Limited media support
- Windows only
Comparison of programs

<table>
<thead>
<tr>
<th></th>
<th>Transcriber</th>
<th>ELAN</th>
<th>Toolbox</th>
<th>FLEX</th>
<th>WeSay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio time-alignment</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Video time-alignment</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Multi-tier annotation</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Interlinear support</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Lexicography</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Word collection</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Simple to learn</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Special char. input</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>XML data</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

What about the non-linguist?

- How can community members (or other interested people) get something out of your work?
- (Maybe they don’t like reading grammars, or dictionaries)
- There are a few ways to allow people to listen to or view your recordings
Some ways to distribute

• People have distributed CDs, or made iTunes libraries out of their archive of recordings of songs, stories etc.
• Others have made DVDs (with or without subtitles) of recorded video
• Now there are also tools for online delivery, for example via YouTube or HTML5 browsers.