

## F4transkript, a simple interface for efficient annotation

Reviewed by CAROLINE JONES and AMIT GERMAN, *Western Sydney University*

**1. Introduction** F4transkript is a software program for annotating audio or video recordings. As argued by Thieberger and Berez (2012: 129), from our annotation tool(s) we want flexibility, interoperability, and archivability. Ideally, we also want to create our corpora as efficiently as possible, in teamwork involving native speakers or other members of the local speech community. F4transkript is a program that can help us achieve these goals.

F4transkript (which we will refer to henceforth by its short name, F4) offers a simple interface with features that also seem likely to promote efficient annotation. F4 is less powerful but also considerably less complex than ELAN (Berez 2007), FLE<sub>x</sub>, or SayMore (Moeller 2014); it complements rather than replaces those tools when it comes to language documentation. We have found F4 to be practical software for non-linguists and linguists alike to create time-aligned annotation in a plain text file format that subsequently can be imported to an ELAN tier. F4 is surprisingly engaging to use and has a quick learning curve. The primary purpose for which we have used F4 is creating a first-pass orthographic transcription but it could equally be used for entering a free translation or other time-aligned commentary.

This review is based on F4transkript version 5.6 for Windows/Linux, which we have found to be slightly better featured than the Mac version, which is called F5transkript (version 2.4). As of February 2016 these are the stable versions of the program but there is a beta 6.0 version available.

F4 software is software for transcription which has been developed and put out for more than ten years by Thorsten Dresing and Thorsten Pehl, researchers at the University of Marburg and co-founders of the company [audiotranskription.de](http://audiotranskription.de). The developers have steadily released a new version every 1–2 years since 2004, and have current plans for further improvements. On our enquiry as to future maintenance plans the developers report “no plans at all” to discontinue the software as it is one of their main products. The latest version is a beta version which uses open source libraries for all media components; the code for GUI components is private.

The native file format for F4 is .rtf with options to output transcripts as .txt, XML or .srt (Subrip subtitle format). In our work we have used the .txt output format only. F4 supports Unicode. The developers do report a problem, found with some other text editors, with opening documents made with a different codepage (for example, opening on a German-configured computer a transcript that was made with a Chinese-configured computer).

In language documentation work, F4 has been used to facilitate annotation by native speakers in the large-scale Chintang Language Research Program in Nepal, headed by Sabine Stoll and Balthasar Bickel at the University of Zurich (e.g. Stoll

& Bickel 2012; see also Chintang Language Research Program, 2016). This inspired our team's use of the software in recent documentation of the north Australian Kriol variety spoken by adults and children at Barunga Community, Northern Territory, Australia.

**2. Using F4** The F4 program has a simple interface that is practical on a small screen such as a laptop without an external monitor. When the program is opened, there are two windows (as shown in Figure 1): an upper panel for entering transcription, and a lower panel that displays the relevant part of the waveform (and in the lowest and thinnest panel, a bird's eye view of position in the waveform) once the media file is loaded.

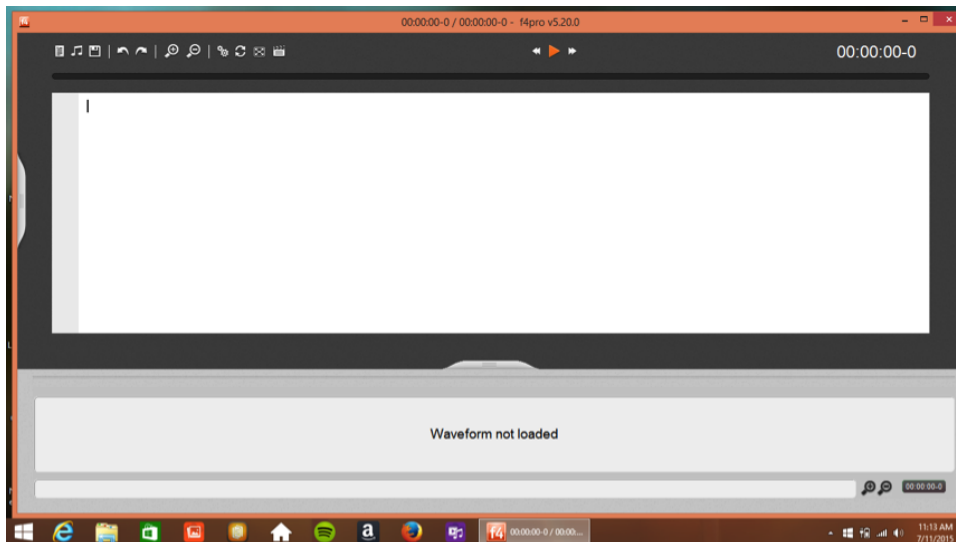


Figure 1. The F4 interface when the program is opened

Clicking on the grey tab to the left of the transcription window opens the control panel (see Figure 2) and clicking on it again conveniently hides it. This saves space on the screen for the main job of transcription.

It is also quite a useful feature that the control panel can be hidden. This means less chance for the user to inadvertently change the settings.

- vary the replay speed e.g., 0.75 (without altering the voice pitch)
- vary the rewind amount (default = 2 sec)
- set up 'hot key' sequences for commonly typed words or names
- set the format of the time-stamp(s)
- specify the location of the time-stamp(s) (segment start and/or end)
- do automatic backup to a local directory e.g., every 5 minutes

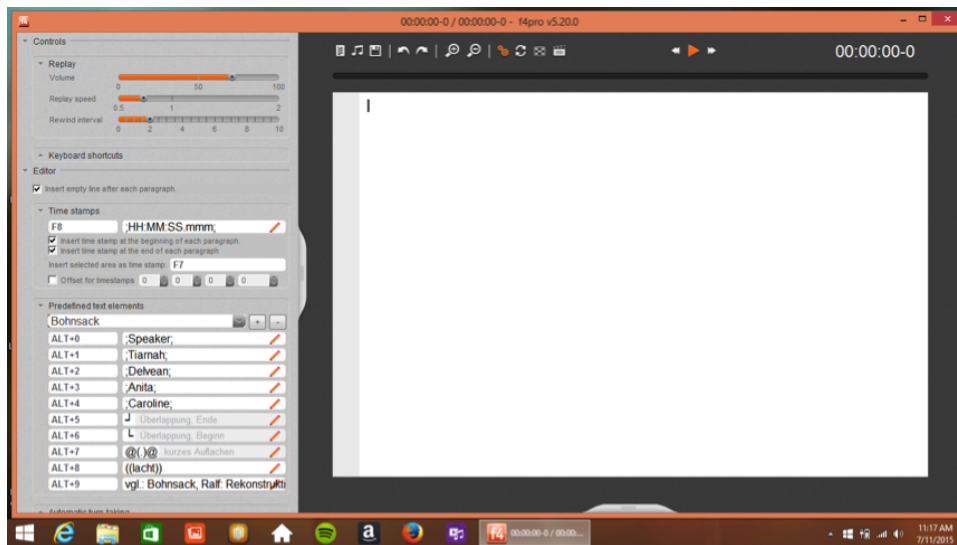


Figure 2. F4 interface with control panel (left) and transcription panel (right)

F4 has been designed to promote faster transcription, primarily by segmenting the media during transcription and avoiding much need to use the mouse or to scroll.<sup>1</sup> F4 also potentially speeds up transcription in that its control panel allows the user to set up hotkeys for frequently typed text, and to reduce the reply speed. The latter is a particular advantage because it tends to mean that fewer replays are required per segment transcribed. F4 can also be used with a foot pedal to free the hands for typing.

F4 is so called because the program co-opts the function F4 key on the keyboard to control playback of the media file. Once the F4 program is installed and the media file is loaded, pressing the F4 key plays the media file, and pressing F4 again stops the media file. At this point, if you enter the text of what you have heard, then press the Enter/Return key, that text is entered along with the time-stamping of the media segment you have just played. This sequence of steps, when repeated for each utterance or turn, creates an annotation that is laid out vertically, like the script of a play. See Figure 3 for an example of a partial transcript.<sup>2</sup>

When you press the F4 key again to resume transcription, there is a default autorewind of 2 seconds which is a useful amount to return to context and hear the last two or three words of recording again before moving on. At any point it is also possible to rewind the media by pressing F3, or to fast forward by pressing F5. Wherever you are in the media file when you press Enter/Return, that is the time point that will be used in the time-stamping. Pressing the F8 key enters the current media file time

<sup>1</sup>To take advantage of these features, it is best to use F4transkript on a PC model on which the function keys are unallocated (such as an Acer, Toshiba or IBM laptop). Otherwise you may need to use a combination keystroke such as Fn+F4 which is slower.

<sup>2</sup>Note that in this mockup recording the speaker is varying between Kriol and English, and the orthography is a rough attempt to keep up with these changes.

(i.e. where the audio has been paused) and is useful in editing or time-stamping the start of a transcript.

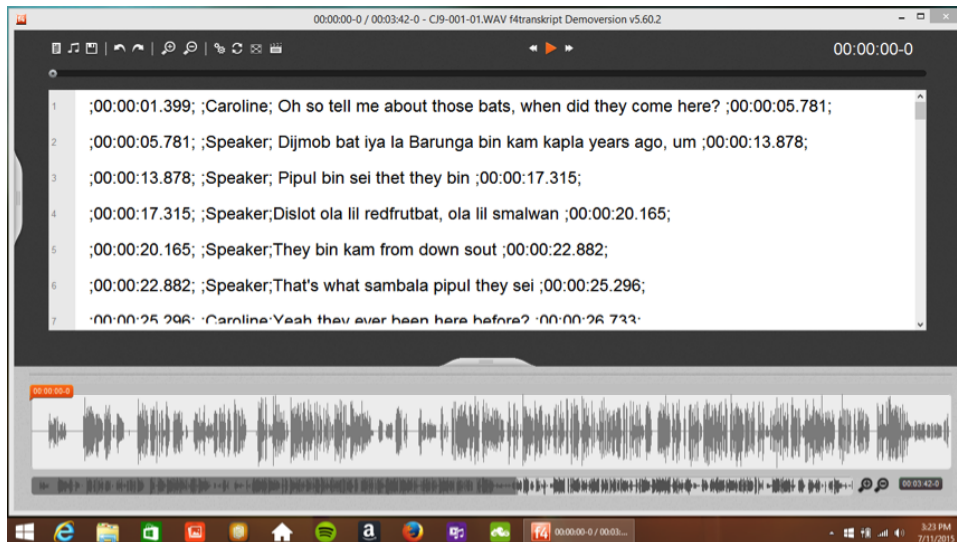


Figure 3. A transcribed conversational text in F4

Note that in Figure 3 there is a time-stamp at the beginning and the end of each transcribed segment. The end of one transcribed segment is marked with the exact same time point as the start of the next transcribed segment. It is not strictly necessary to use F4 to annotate without gaps in this way but it makes for faster annotation than segmenting precisely. (Depending on the goals of analysis or end-product presentation, a closer alignment to the signal is possible at a later stage in corpus creation using a forced alignment tool such as WebMAUS. See Schiel (1999), and Strunk, Schiel and Seifart (2014) for details.)

**3. Settings in F4 that facilitate import to ELAN** F4transkript is a useful program for relatively quick entry of one type of time-aligned information. We have used it for entry of a first-pass orthographic transcription. The initial orthographic transcription can be easily imported into one or more tiers in ELAN if the following settings are used in F4 software before starting transcription.

*Time-stamps:* In the control panel, in the 'Time stamps' drop down, ensure time-stamps are set for the beginning and end of each segment transcribed, with the format:

;HH:MM:SS.mmm;

This time-code format sets up the semi-colon to be the delimiter on import into ELAN. The decimal point between seconds and milliseconds is an ELAN-compatible time format.

*Tier names:* In the control panel in F4, set up hotkey sequences with whatever you want the resulting tier name to be in ELAN. In our conversational recordings we set up one hotkey per speaker in the session, e.g.:

Alt-o ;Speaker;  
Alt-1 ;Caroline;

Again, the semi-colon acts as delimiter on import to ELAN.

*Save as .txt:* By default the transcript will be saved as .rtf. This is potentially useful as it can display different speakers in different color fonts, for example. In our work we have selected the output format as .txt (from within the F4 program itself) to avoid complexities in opening file formats in the longer term, and since the .txt file imports well into ELAN.

*Import to ELAN:* Using a text editor, first do two things: (1) Replace any instances of two semi-colons with one semi-colon. This avoids issues with the ELAN GUI import accepting a limited number of columns. (2) Manually delete the second instance of the final time-stamp (which has been automatically introduced by the control panel selection to time-stamp at start and end of each segment). Launch ELAN, and import the tab-delimited .txt file using the File > Import menu. Select the columns and specify the data type in each column (start time, speaker name as tier name, annotation, and end time). Specify the semi-colon as the delimiter, and the transcript will import to the relevant tier(s) in ELAN, with one tier per tier name you used in F4.

**4. Preventing or resolving issues** One of the strengths of F4 is that it is easy for a non-linguist or non-technical person to use it to annotate. Nevertheless we have found a few ways in which transcription can go awry using F4.

In our work we have opted to always save the transcript as a plain .txt file. The default output file format in F4, however, is .rtf. If the user instead saves the annotation in F4 as an .rtf file, it is a simple matter to open it in a text editor and do a Save As to .txt. It is also possible to just re-save within F4 as a .txt file, but it would be necessary to delete the .rtf copy to avoid confusion in ongoing transcription.

F4 requires only fairly basic computer literacy, but there are potential pitfalls for beginner computer users. If the user does not ensure that the cursor is placed at the end of the annotation (e.g. when they resume work after a break), then transcription can get entered into the middle of existing annotation. If this happens the time-stamps will be out of order. To fix this, you can cut and paste the segment that is out of time alignment, or just backspace and redo the problematic part, inserting time-stamps as necessary using Enter/Return and F8 keys.

It is important that the user learns early that pressing the Enter/Return key enters the time-code. It needs to become automatic that they press Enter/Return *after* the transcription is keyed in for the utterance/turn they have just played. If instead the Enter/Return key is pressed after the user has listened to the utterance/turn and before the transcription is keyed in, the result is that the transcription is out of time-

alignment, usually delayed by around four seconds or more (depending on the length of the segment).

To support novice transcribers or someone who is underconfident with computers, it can be useful to make a ‘starter’ annotation file for each media file, containing just the first utterance or two already transcribed. That way they do not have to handle the initial step of starting a transcript. Instead they can just continue with transcription, using instructions like those below, for example:

===

*How to use F4 on an existing transcript*

1. Double-click the F4 icon on the desktop.
2. Click the document icon to Open a Transcript. (It’s important to start by getting the transcript file, not the sound file! Otherwise it won’t work.)
3. Browse to the text file you want, then click it, then click Open.
4. Then when you’re asked to choose the media file, browse to the audio file, click it, then click Open.
5. Press F4 to play a bit, then F4 to stop. Go until the cursor (in the audio window) is where you want to be in the recording. You can also use F3 to rewind / go back. Use F5 to fast forward.
6. When you are in the right spot in the media file, make sure the cursor (in the transcription window) is at the end of the transcript.
7. Press F4 to play what someone says, then F4 to stop when you are ready to transcribe. Keep what they’ve said in your memory!
8. First just type the ALT+1/2/3/O for the person who is talking. (Their name will be filled in with semi-colons on either side. Leave all the punctuation in!)

ALT+1 = Jaidine  
 ALT+2 = Speaker (e.g. the child)  
 ALT+3 = Lee

9. Then type in what the person said. Then press the Enter key once. The timecode is automatically inserted.
10. Go back to Step 7 above, i.e. type F4 and go on to the next bit of talking. You’ll notice that the audio backtracks 2 seconds each time so that you can keep track of where you’re up to in the audio.
11. Click on the Save icon (the floppy disk). (Don’t click on Save Transcript or Save Transcript as). The transcript is saved as a .txt (text) file.

===

**5. Conclusion** F4 is annotation software with a simple interface. F4 can be used as one step in the creation of a language documentation corpus, as a complement to more technical tools such as ELAN. This review has focused on using F4 as a tool for basic orthographic transcription by linguists and non-linguists alike. F4 is potentially also useful for involving community members in writing a time-aligned translation or other commentary on a recording. This review has focused on the Windows/Linux version of the software (F4transkript) which we have successfully used on cheap, low-spec PC laptops. The Mac version, F5transkript, is less fully featured. Unfortunately, for example, it does not display the audio waveform.<sup>3</sup>

F4 has a simple interface, which is easy to learn and well sized for use on a laptop in the field. F4 also has built-in features that should facilitate efficient transcription. While we have not run a strict comparison of suitably experienced users of F4transkript and another current program (e.g., ELAN, Simple ELAN), we have been pleased with the speed of transcription using F4transkript. Native speakers of Barunga Kriol on our team achieve on average a rate of 4:1 (ratio of transcription time to actual media time) when transcribing adult speakers. When transcribing child speakers, the average rate is around 7:1. They are not touch typists and for simplicity do not currently use a footpedal (although a footpedal is a plug-in option with F4). The rates of 4:1 for transcribing adult speakers and 7:1 for transcribing child speakers seem efficient, and compare well with published estimates of transcription time. For excerpts from a naturalistic corpus of adult-infant interaction, Roy and Roy (2009) report ratios of transcription time to actual media time (for English, by native speakers) of approximately 9:1 for CLAN in sonic mode, and 6:1 for Transcriber, and these figures do not include the time taken for annotating speaker identity/turn (which is already included in the rates we report for F4transkript above). Other published estimates for initial transcription rate are ratios of between 10:1 and 20:1 for English (Tomasello & Stahl 2004:104) and up to 35:1 for smaller endangered languages (Auer et al. 2010:31).

The output of F4 is either a .txt or .rtf file. By choosing to save as .txt and by setting the hotkey and time-stamp options in the control panel (as described above in Section 4), it is straightforward to import the F4 annotation to one or more tiers in an ELAN. If there is more than one speaker it is straightforward to assign each speaker's transcription in F4 to its own ELAN tier. One major limitation with the way F4 is set up to be used most easily is that overlapping speech of two talkers is segmented sequentially, not as accurately overlapping intervals. It should be noted that it is technically possible to transcribe accurate overlaps using F4 but it would considerably slow down the rate of transcription to do so. Whether this is an issue might depend on the goals of the analysis or representation of the recording.

<sup>3</sup>For detailed instructions on using the Mac version, F5transkript, as well as F4transkript, see the course note PDFs from the course "Making Friends with Your Corpus Data" from the 1st Summer School of the Australian Research Council Centre of Excellence for the Dynamics of Language, December 3–7 2015, Sydney, Australia, at: <http://www.dynamicsoflanguage.edu.au/education-and-outreach/train-with-us/summer-school-2015/meet-the-course-instructors/>

<b>Primary focus:</b>	Entering time-aligned annotation for audio or video.
<b>Pros:</b>	Quick to learn, simple annotation interface, variable playback speed, can set up hot keys for frequent text, can use with a foot pedal, creates a text file (e.g., .txt) importable into e.g., ELAN, has Unicode support. F4 does not require much processor speed or memory; it runs well on a 2015 base-model PC laptop and on older machines like mid-range PCs from the mid-2000s running Windows XP or newer. Codec packs are available for download for PC and Mac if media files are not being played properly from within the software: <a href="http://www.audiotranskription.de/english/f4.htm">http://www.audiotranskription.de/english/f4.htm</a>
<b>Cons:</b>	If transcribing more than one speaker in a conversation in which speakers overlap, F4 makes it easy to segment sequentially, but it is more time-consuming to segment as accurately overlapping intervals. In the Mac version (F5transkript) the audio waveform is not displayed, the control panel cannot be as easily hidden as in the Windows version, and only one time-stamp per segment is possible. (These are enough disadvantages for us not to have used F5transkript ourselves.)
<b>Platforms:</b>	Windows 7, 8, or 10 (& Windows Media Player); Linux Ubuntu 14.04, 15.10 & VLC Player; Mac
<b>Open source:</b>	The new beta version uses open source libraries for media components, but the code for GUI components is proprietary. The software is priced at 69,90 Euros; educational and multi-user pricing is available. As software updates come out it is currently possible to download, install and activate at no additional cost, using the originally purchased license key.
<b>Reviewed Version:</b>	F4transkript version 5.6 (9 June 2015) for Windows/Linux, and F5transkript version 2.4 (131) (2015) for Mac.
<b>Application Size:</b>	3.0 MB
<b>Documentation:</b>	Manual available in English and German from: <a href="https://www.audiotranskription.de/english/downloads-en.html">https://www.audiotranskription.de/english/downloads-en.html</a>
<b>Download from:</b>	<a href="https://www.audiotranskription.de/english/download-en.html">https://www.audiotranskription.de/english/download-en.html</a>

### References

- Auer, Eric, Peter Wittenburg, Hans Sloetjes, Oliver Schreer, Stefano Masneri, Daniel Schneider & Sebastian Tschöpel. 2010. Automatic annotation of media field recordings. In Caroline Sporleder & Kalliopi Zervanou (eds.), *Proceedings of the ECAI 2010 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2010)*. Lisbon, 2010. 31–54. <http://ilk.uvt.nl/LaTeCH2010/LPF/ws16.pdf>.
- Berez, Andrea L. 2007. EUDICO Linguistic Annotator (ELAN) from Max Planck Institute for Psycholinguistics. *Language Documentation & Conservation* 1. 283–9.
- Chintang Language Research Program. 2016. *The Chintang Language Research Program*. <http://www.clrp.uzh.ch/>.
- Dingemans, Mark, Jeremy Hammond, Herman Stehouwer, Aarthi Somasundaram & Sebastian Drude. 2012. A high speed transcription interface for annotating primary



- linguistic data. In *Proceedings of the 6th EACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2012)*. Avignon, France, April 2012. 7–12. <http://dl.acm.org/citation.cfm?id=2390359>.
- Moeller, Sarah R. 2014. SayMore, a tool for Language Documentation Productivity. *Language Documentation & Conservation* 8. 66–74. <http://hdl.handle.net/10125/4610>.
- Roy, Brandon C. & Deb Roy. 2009. Fast transcription of unstructured audio recordings. In Uther, Maria, Roger Moore & Stephen Cox (eds.), *Proceedings of the 10th International Conference of the International Speech Communication Association (Interspeech 2009)*. Brighton UK, September 2009. 1647–50. [http://www.isca-speech.org/archive/archive\\_papers/interspeech\\_2009/papers/i09\\_1647.pdf](http://www.isca-speech.org/archive/archive_papers/interspeech_2009/papers/i09_1647.pdf)
- Schiel, Florian. 1999. Automatic Phonetic Transcription of Non-Prompted Speech. In Ohala, John J., Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.), *Proceedings of ICPhS 1999*. San Francisco, August 1999. 607–610. [https://www.phonetik.uni-muenchen.de/forschung/publikationen/ICPhS99\\_Schiel.pdf](https://www.phonetik.uni-muenchen.de/forschung/publikationen/ICPhS99_Schiel.pdf)
- Stoll, Sabine & Balthasar Bickel. 2012. How to measure frequency? Different ways of counting ergatives in Chintang (Tibeto-Burman, Nepal) and their implications. In Seifart, Frank, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts & Paul Trilsbeek (eds.), *Language Documentation & Conservation Special Publication No. 3 (August 2012), Potentials of Language Documentation: Methods, Analyses, and Utilization*, 83–89. <http://nflrc.hawaii.edu/lcd/spo3/>.
- Strunk, Jan, Florian Schiel & Frank Seifart. 2014. Untrained Forced Alignment of Transcriptions and Audio for Language Documentation Corpora using WebMAUS. In Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC14)*, 3940–3947. Paris: European Language Resources Association. <http://www.lrec-conf.org/proceedings/lrec2014/index.html>.
- Thieberger, Nicholas & Andrea L. Berez. 2012. Linguistic Data Management. In Thieberger, Nicholas (ed.), *The Oxford Handbook of Linguistic Fieldwork*, 90–118. Oxford: Oxford University Press.
- Tomasello, Michael & Daniel Stahl. 2004. Sampling children's spontaneous speech: how much is enough? *Journal of Child Language* 31. 101–21.

Caroline Jones  
 caroline.jones@westernsydney.edu.au

Amit German  
 A.German@westernsydney.edu.au