

in the corpus. However, the corpus also includes coding in angle brackets for, say, syntactic roles (e.g. ⟨sub⟩ for subject, ⟨prd⟩ for predicate), which you need to exclude from the text count. Using an ordinary find-and-replace tool will not locate all of these items at once (you would need to search for each of them individually). With a regular expression search you can locate all patterns of the form ‘three characters enclosed in angle brackets’ with the regular expression of the form `<...>`, where a full stop or period represents ‘any character except return’, thus looking for three such characters inside angle brackets.

MS Word has a very weak form of regular expression search. There are a number of tools with fully-featured regular expression functions, including OpenOffice.org, Text Wrangler (Mac), EditPad Pro (Windows), and regexxer (Linux). Regular expressions can swap the order of items in a text, and insert and change regular patterns based on shape rather than content.

4.3.5 Annotation of media: time alignment and interlinearization

Transcription and translation of recordings provides another opportunity to create well-structured data. A time-aligned transcription matches sections of media to a textual representation of the contents of the recording, creating links between the two via time codes. There are several advantages to aligning text to media over creating an unaligned transcript (i.e. a transcript created in a word processor). From a research standpoint, time alignment allows access to language as an audiovisual signal, rather than just a written interpretation. You will be able to quickly search and retrieve particular portions of the media to assist you in developing an analysis. Most time alignment software produces an XML file that is both archivable and human-readable, allowing the link between the recording and the transcript to endure beyond the life of the software used to create it. From a presentation standpoint, a time-aligned transcript can be quickly and easily converted into display formats via tools like CuPED,²⁹ Eopas,³⁰ and Annex/Trova.³¹

p. 113 There are several software options available for time alignment designed for documentary linguistic use, including ELAN,³² Transcriber,³³ and EXMARaLDA.³⁴ When choosing software, things to look for are:

- (i) flexibility—does the tool support multiple speakers, overlapping speech, and hierarchical relationships between tiers for interlinearization, comments, or annotation of gesture?
- (ii) interoperability—do the tool's import and export options allow easy conversion between the other tools in your workflow, like a lexical database or a text editor?
- (iii) archivability—does the tool produce a well-documented XML file?³⁵