

3-6-2018

Audiovisual Media Annotation Using Qualitative Data Analysis Software: A Comparative Analysis

Liliana Melgar Estrada
University of Amsterdam, melgar@uva.nl

Marijn Koolen
Huygens ING, marijn.koolen@huygens.knaw.nl

Follow this and additional works at: <https://nsuworks.nova.edu/tqr>



Part of the [Library and Information Science Commons](#), and the [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#)

Recommended APA Citation

Melgar Estrada, L., & Koolen, M. (2018). Audiovisual Media Annotation Using Qualitative Data Analysis Software: A Comparative Analysis. *The Qualitative Report*, 23(13), 40-60. <https://doi.org/10.46743/2160-3715/2018.3035>

This Article is brought to you for free and open access by the The Qualitative Report at NSUWorks. It has been accepted for inclusion in The Qualitative Report by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.



Audiovisual Media Annotation Using Qualitative Data Analysis Software: A Comparative Analysis

Abstract

The variety of specialized tools designed to facilitate analysis of audio-visual (AV) media are useful not only to media scholars and oral historians but to other researchers as well. Both Qualitative Data Analysis Software (QDAS) packages and dedicated systems created for specific disciplines, such as linguistics, can be used for this purpose. Software proliferation challenges researchers to make informed choices about which package will be most useful for their project. This paper aims to present an information science perspective of the scholarly use of tools in qualitative research of audio-visual sources. It provides a baseline of affordances based on functionalities with the goal of making the types of research tasks that they support more explicit (e.g., transcribing, segmenting, coding, linking, and commenting on data). We look closely at how these functionalities relate to each other, and at how system design influences research tasks.

Keywords

QDA Software, QDAS, CAQDAS, Qualitative Data Analysis, Audiovisual Data, Media Scholars, Research Tasks, Interoperability, Data Models

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Acknowledgements

We would like to thank our reviewers for their valuable input.

Audiovisual Media Annotation Using Qualitative Data Analysis Software: A Comparative Analysis

Liliana Melgar-Estrada

University of Amsterdam / Netherlands Institute for Sound and Vision, The Netherlands

Marijn Koolen

Huygens ING, Royal Netherlands Academy of Arts and Sciences, The Netherlands

The variety of specialized tools designed to facilitate analysis of audio-visual (AV) media are useful not only to media scholars and oral historians but to other researchers as well. Both Qualitative Data Analysis Software (QDAS) packages and dedicated systems created for specific disciplines, such as linguistics, can be used for this purpose. Software proliferation challenges researchers to make informed choices about which package will be most useful for their project. This paper aims to present an information science perspective of the scholarly use of tools in qualitative research of audio-visual sources. It provides a baseline of affordances based on functionalities with the goal of making the types of research tasks that they support more explicit (e.g., transcribing, segmenting, coding, linking, and commenting on data). We look closely at how these functionalities relate to each other, and at how system design influences research tasks. Keywords: QDA Software, QDAS, CAQDAS, Qualitative Data Analysis, Audiovisual Data, Media Scholars, Research Tasks, Interoperability, Data Models

Introduction

The so-called digital turn (Desrochers & Apollon, 2014) has generated digital content at an unprecedented pace and continues to transform research practices in all disciplines at many levels. This digital transition has originated both from the rising availability of digitized or digitally born sources and publications, and also from a wide range of information processing “tools”. Indeed, these tools not only assist scholars in performing traditional tasks more efficiently, but also challenge them to reflect on their methodology and methods, as they increase the possibilities for creating, collecting, analyzing and visualizing source materials on both small and large scales.¹ When these tools are used by groups, or are created as web applications, they help to expand collaborative analyses and knowledge sharing.

Specialized tools for qualitative analysis are being used in academic research more often. The DiRT Directory², a registry of digital research tools for scholarly use, listed more than four hundred fifty tools when last updated in 2015. These include systems for capturing, creating, enriching, analyzing, storing or disseminating digital content. Among these, a group of tools, namely Qualitative Data Analysis Software (QDAS) has existed for more than thirty years (Silver & Patashnick, 2011), and its number is increasing.³ Given the variety of tools, researchers need clear criteria for selecting the most appropriate one for the task. Ideally,

¹ See, for instance, Berry (2012) for a discussion about the epistemological impact of digital technologies in the humanities, and the move towards the “computational turn.”

² <http://dirtdirectory.org/>

³ See, for instance, the list provided by the Social Science Software inventory, SoSciSo (Gey, n.d.).

software developers would provide users with the means to combine, move and store the outputs of their computer-assisted analyses across and outside of specific software packages.

Audio-visual (AV) materials have traditionally constituted only a small part of the sources studied by humanities scholars and social scientists (Benardou et al., 2013). Now, though, these materials have been introduced in traditionally text-based disciplines (Clivaz, 2016; Silver & Patashnick, 2011), and their use continues to grow among the disciplines that make them the object of study (e.g., film, television and media studies or visual anthropology).⁴ The use of AV sources (which broadly encompass film, television, radio, sound recordings, or any other form that combines image and sound)⁵ complexifies the process of making informed software choices for two reasons. First, despite automatic indexing of AV sources advancing rapidly (Huurnink et al., 2012; Weigel, 2016), to a certain extent they still constitute a “blind medium” for retrieval (Sandom & Enser, 2001). Unlike text, these sources usually require manual sequential viewing and annotation, in order to transcode the content (e.g., creating a transcription), or to identify meaningful units at different levels, such as objects or actions, spoken words, or abstract ideas.

In the case of text, search tools or natural language processing techniques can more readily provide indicators of recurring words, or even help identify abstract concepts during preliminary analysis. Second, AV sources are rarely used as the sole source of data, and thus, contextualization, via textual material or other media is also required. Thus, tools should support not only data analysis but also data preparation, manual annotation, and the use of multiple media types (Clivaz, 2016). Since researchers these days are actively using software to assist in these tasks, a greater understanding their affordances and how their use impacts the analysis process and its outcome is needed.

This paper compares proposes criteria to guide scholars in evaluating how these tools can support their research. To this end, we compared two categories of software: QDAS packages (NVivo for Mac 11.3.2 and Transana 3.01) and dedicated AV analysis software (ELAN for Mac OS 4.9.4). A further aim of this comparison is to contribute to the ongoing discussion, both in the information science domain and in the scholarly community, about software interoperability (see Evers, this issue). The work presented in this article has been done in the context of CLARIAH, a national digital research infrastructure project for arts, humanities and social sciences in The Netherlands,⁶ in which scholars and information specialist work together to facilitate access to cultural heritage collections and data in a sustainable way, by developing a series of open source and interoperable tools.

Research Tasks in Qualitative Audio-Visual Analysis

While each discipline may approach qualitative research in a specific way, scholars have identified a core set of frequently used tasks⁷ (Tesch, 1990). Key to this idea is the notion of so called “primitives,” which Unsworth (2000) named and defined as “some basic functions common to scholarly activity across disciplines, over time, and independent of theoretical orientation.” Some of these primitives are: discovering, annotating, comparing, referring,

⁴ For more details about the use of visual sources in other disciplines see: Schnettler and Raab (2008), Noordegraaf (2016), and Stanczak (2007).

⁵ More complete definitions and perspectives of the term “audio-visual” are in Hewett and Barber (2013), Usai et al. (2008), UNESCO (2012).

⁶ The CLARIAH project (<https://www.clariah.nl/>) is a Dutch national initiative to build an infrastructure for digital humanities research, part of the pan-European initiatives DARIAH (Digital Research Infrastructure for the Arts and Humanities) and CLARIN (Common Language Resources and Technology Infrastructure).

⁷ Even though the concept of task is not clearly defined, it is often used in Information science or Human Computer Interaction domains to refer to a series of steps or activities which are logically organized to achieve a goal (e.g., Ingwersen & Järvelin, 2005, p. 73).

sampling, illustrating, representing (Unsworth, 2000). Primitives have been associated with information-related work (Palmer et al., 2009) or with concrete tasks performed through data capturing, creation, enrichment, analysis, interpretation, storage and dissemination (Borek et al., 2016). Research in the field of information behavior has also indicated that there may be common, though not necessarily sequential, stages in the research process used by scholars in similar domains (e.g., literary scholars: Chu, 1999, or media scholars: Bron et al., 2015), in which analysis is one of the key phases (see for instance Kendall, 2012).

While the idea of common phases of research across disciplines may be debatable, scholars seem to agree that *analysis*⁸ is part of qualitative research. Because QDAS may be used for this part of the study (Woods et al., 2016), the identification of the main tasks performed by scholars utilizing AV media during analysis work, with or without software, becomes central to our understanding of this kind of work. The identification of these tasks can inform improvement of QDAS or any other information system for supporting scholarly work.

Task analysis is an approach for understanding software support in the analysis of AV media. It assumes that research is constituted by a sequence of tasks (either conceptual or “mechanical”)⁹ that are reflected in a tool’s features. In the next sections, we describe the task analysis used to compare three software packages. It is based on the tasks defined by Silver and Patashnick (2011), Silver et al. (2011), and Melgar et al. (in press). We identify *transcribing*, *segmenting*, *coding*, *linking* and *commenting* as a set of fundamental common core tasks of AV data analysis across disciplines. We have categorized these five tasks under the umbrella term “scholarly annotation” for AV-media centered research.¹⁰

Transcribing

Converting the audio-visual signal into textual, natural language representations can be done in a variety of ways. Evers (2011) described four formats: pragmatic verbatim, gisted, Jeffersonian and Goodwinian, and we propose two additional types: translations, and descriptions. Transcribing and describing are essential to AV analysis because the message cannot otherwise be accessed directly. Transcribing is an analytic task, since it attempts to “transcode” or capture elements from the AV message into textual forms that can be more easily manipulated. Transcription tasks (conversion of speech into text) can be done by hand (e.g., listening and manual typing), automatically (also known as “automatic speech recognition”), or semi-automatically (a combination of the two).

Segmenting

Because of the limitations of the human mind in processing large amounts of content, Tesch (1990) claimed “the analyst concentrates on sets of smaller and more homogeneous chunks of material at any one time” (p. 96). Segmenting (called “marking” by Tesch) is an essential scholarly analytic task, since determining the units of analysis (the fragments or portions to focus upon) lays the foundation for subsequent synthesis and interpretation. Segmenting tasks can be done manually, automatically (e.g., by using shot-boundary detection,¹¹ or automatic audio recognizers), or a combination of both.

⁸ Analysis (i.e., separation) should not be confused with *synthesis* (or *interpretation* using the terms of the taxonomy referred in Borek et al., 2016, and by Marsden et al., 2007).

⁹ Silver and Shelly (1995, as cited in Silver & Lewins, 2014) distinguish between “conceptual” tasks (e.g., reading, questioning, categorizing, etc.), and “mechanical” tasks (e.g., storing, organizing, retrieving data), explaining how mechanical tasks are the grounding of the conceptual tasks.

¹⁰ See note (k) under Table 2.

¹¹ State of the art techniques for automatic video processing are summarized in Weigel (2016).

Coding

Scholars enrich their analytic units (fragments) with their annotations, which, depending on the data analysis approach, can take the form of codes (i.e., short keyphrases or summative terms), more open natural language representations, or comments. This is done during the entire research process in a cyclical, non-linear way (Evers, 2015; Saldaña, 2016).

Linking

Using the taxonomy proposed in Borek et al. (2016)¹², linking refers to the tasks of creating associations between objects of investigation with the purpose of interpretation. Linking is essential to contextualization tasks, which aim to find transtextual relations between the object of analysis and other “texts” (Genette, 1997). This task occurs, for example, during analytical segmenting or coding tasks, by creating links (“hyper-linking”) to external or internal objects within a research project.

Commenting

Besides creating links, another aspect of contextualizing is commenting. This consists of using analytic memos as an essential part of the analysis process, to register observations and documents the reasons why certain codes, hyperlinks, or other annotations have been created. *Commenting* tasks are not be confused with “annotating,” since comments are only one form of annotation (see Table 2).

This section introduced five common analytical tasks performed in qualitative audio-visual analysis. The next section presents a typology of software programs that support scholars during these tasks.

Types of Software for Audiovisual Data Analysis

Software packages support researchers in gathering, preparing and analyzing data, and transforming analogue resources and analysis steps into discrete, categorized units. Little is known about how scholars use AV materials or software in their analysis. This may be because the analysis process is not described extensively, or the use of software is not reported (Silver & Patashnick, 2011). However, the modeling choices made when designing software affect both the data itself, the concrete analysis tasks that can be performed, and the order in which scholars can perform them. Thus, a better understanding of the affordances of these tools is warranted.

Based on our inventory of existing tools for video annotation (Melgar et al., in press), we identified three categories of tools: (1) common QDAS packages (QDAS-C), (2) AV media specific QDAS (QDAS-AV), and (3) specialized audio-visual annotation tools (PVA). We selected one tool from each category for our comparison. To identify relevant QDAS-C tools we reviewed existing reviews and inventories, including those created by the University of Surrey (Silver et al., 2011), Gibbs (2014), the Social Science Software inventory (SoSciSo), Silver and Lewis (2014), and Evers et al., 2011. For identifying software geared towards analysis and annotation of AV data (QDAS-AV and PVA) we took into account an inventory conducted by Melgar et al. (in press). One package from each category was selected for review (Table 1) based on these criteria: ability to work with AV media, inclusion in multiple inventories,

¹² The taxonomy of research tasks TaDiRAH (<http://tadirah.dariah.eu/>), an initiative started by DARIAH, and DiRT.

preference for free or open source software (for this reason we chose ELAN above ANVIL), and actively maintained (for this reason we excluded DRS¹³). Three software packages from the QDAS-C category scored equally on these criteria (i.e., ATLAS.ti, MAXQDA, and NVivo). We selected the latter to benefit from experience gained in a previous study (Melgar et al., in press). We excluded from analysis professional video editors and video retrieval prototypes.¹⁴

Table 1. Software packages selected for the analysis

Tool	Type	Survey/Source	Free	Open source¹⁵	Actively maintained¹⁶
Nvivo for Mac 11.3.2	QDAS-C	Evers et al., 2011; Silver et al., 2011; Gibbs, 2014; Silver & Lewins, 2014; University of Surrey, 2014	No	No	Yes
Transana 3.01	QDAS-AV	Marsden et al., 2007; Evers et al., 2011; Silver et al., 2011; Gibbs, 2014; Silver & Lewins, 2014	No	No	Yes
ELAN for Mac 4.9.4	PVA OS	Melgar et al., in press	Yes	Yes	Yes

Next, we will describe the different types of tools and software packages in more detail.

Common QDAS (NVivo)

Common QDAS tools, also known as Computer Assisted Qualitative Data Analysis (CAQDAS) software packages, offer features that correspond to qualitative analysis principles, for instance, from grounded theory (Pickard & Childs, 2013) to discourse analysis (Paulus & Lester, 2016). They were originally developed by social sciences scholars (Silver & Lewins, 2014). Even though QDAS packages were mostly designed for textual sources, most of them now include the ability to analyze AV sources.¹⁷

The package we review here, NVivo, was created in 1981, then called NUD*IST ('Non-Numerical Unstructured Data Indexing Searching and Theorizing'). The main NVivo interface (Figure 1) is arranged in different horizontal panels, which are, starting from the top: (1) ribbon with tools and commands; (2) navigation panel including sources, analytic units, and

¹³ The source code is still openly available.

¹⁴ Examples of these tools are included in Dasiopoulou et al. (2011) or Nixon and Troncy (2014).

¹⁵ Open source means that the source code of the software is open, available to anyone who wants to use or improve it.

¹⁶ We considered the last year of update (i.e., 2015 onwards).

¹⁷ See Gibson et al. (2005) for a historical note.

source information; (3) detailed contents view; (4) video, wave form, and coding stripes along timeline; and (5) transcript table and coding stripes along transcript.

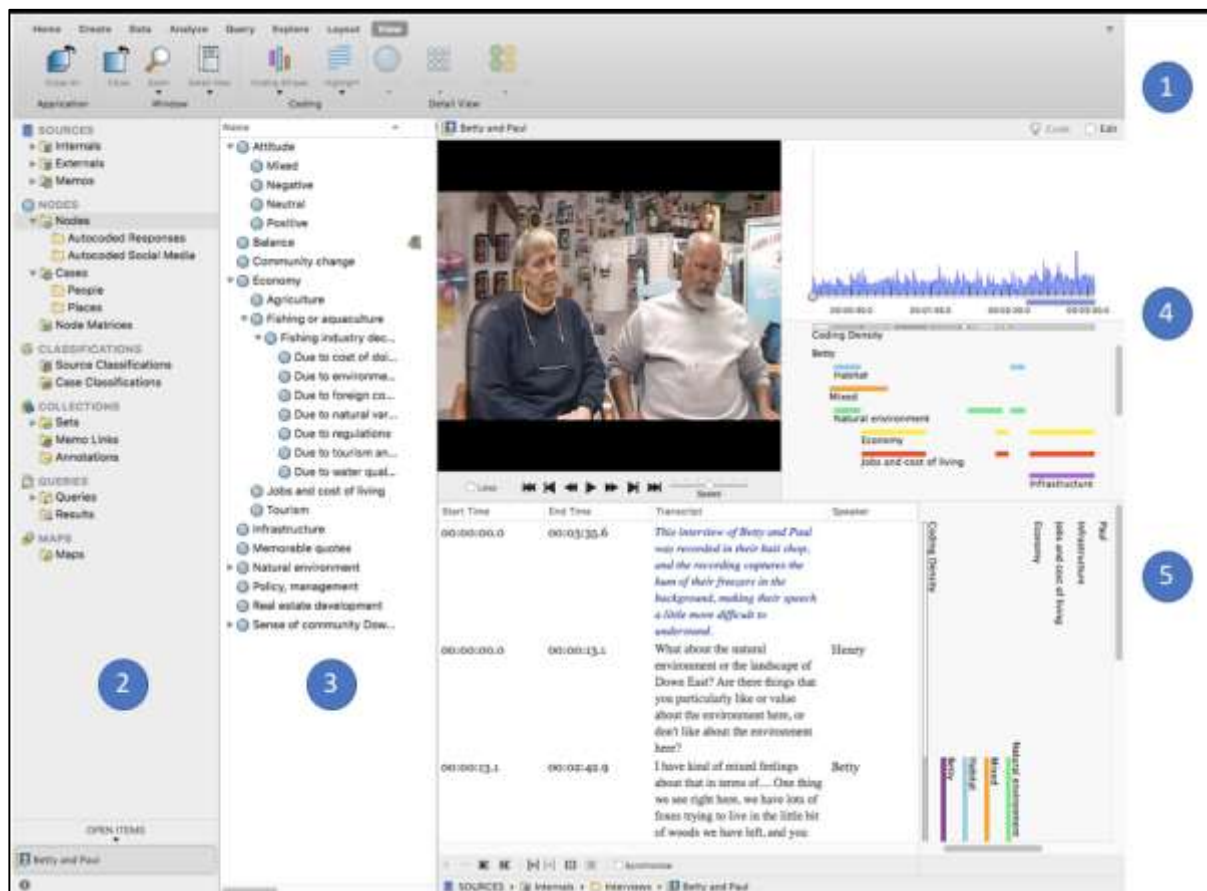


Figure 1. NVivo interface

Audio-Visual Media Specific QDAS (Transana)

Silver and Patashnick (2011) and Leujeune (2013) identified a number of software packages that specialize in the analysis of AV media, such as Mixed Media Grid (MiMeG), Digital Replay System (DRS), Transcriber, Videograph, and Transana.

Our selected package, Transana, was originally created by Chris Fassnacht, released in 2001, and as of 2017 is maintained by David K. Woods. This software specializes in supporting researchers in the transcription and analysis of video, audio, and still images. The main interface (Figure 2) presents: 1) the visualization window with coding; 2) a media window with multi-stream facilities; 3) a document window with multiple simultaneous transcripts; and 4) the data window with main objects, including clips.

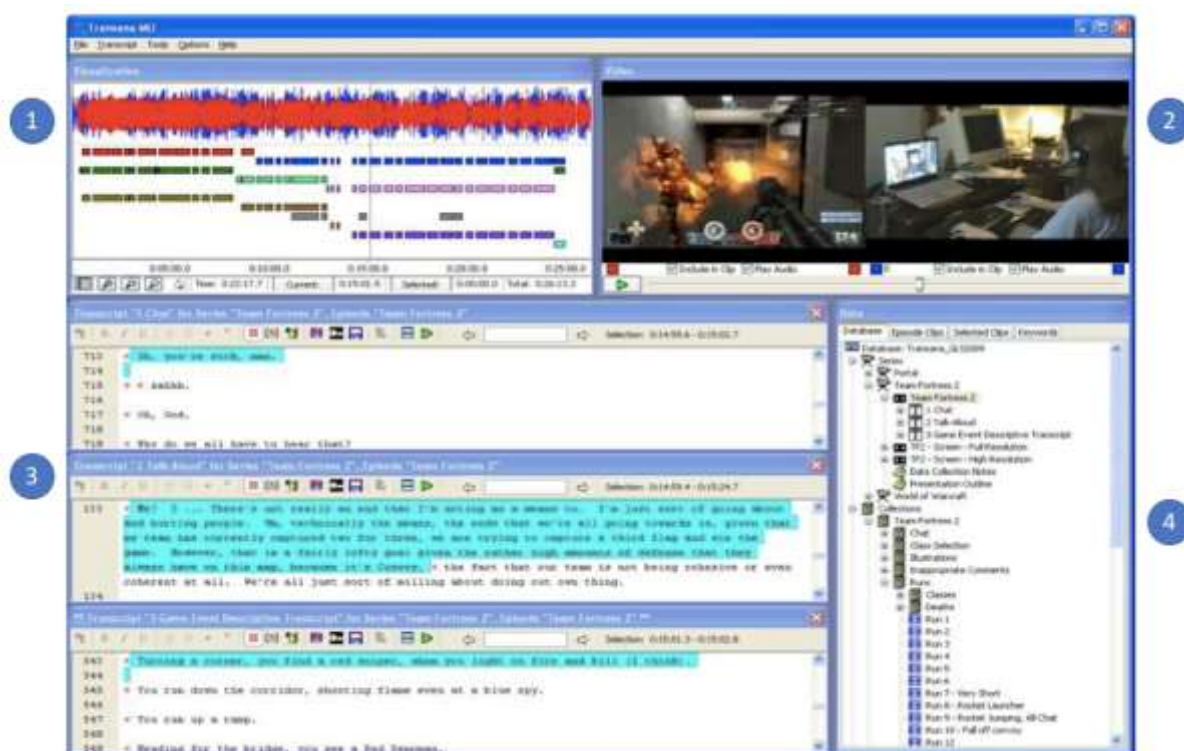


Figure 2. Transana interface

Specialized Audio-Visual Annotation Tools (ELAN)

In the final category are professional video annotation tools tailored to the analytic concerns of, for instance, linguistics, educators, and psychologists. Examples of specialized tools for time-based annotation¹⁸ include EUDICO Linguistic Annotator (ELAN), Anvil, and EXMARaLDA. From the behavioral sciences, examples include Observer and Interact, both of which are proprietary software for mixed methods analyses of multimedia data. In the media studies domain, initiatives include Annotate Digital Video, Exchange on the Net (Advene), Digital Cinema Project (DCP) (Giunti, 2014), and Lignes du Temp, the latter developed by L'Institut de recherche et d'innovation in France, as well as Recall, which originated in the domain of performative arts, specifically for dance analyses. Even though these analytic tools are developed with specific analytic interests, they are also useful to researchers from other disciplines using similar methodological approaches.

The selected package within this group, ELAN, was created at The Language Archive of the Max Planck Institute for Psycholinguistics (Nijmegen, The Netherlands) in the late 1990s. It was developed by and for a community of linguists and communication scholars, and is currently used in many other domains.¹⁹ Its main interface elements are labeled in Figure 3, with annotation tiers in the bottom left (a tier represents a single layer of annotations connected to the timeline), the annotation boxes in the timeline viewer in the center (each box is an annotation on a segment of the timeline), and recognizers in the upper part (recognizers are installable plugins that can automatically detect and annotate certain events, such as shot boundaries, utterances, pauses and turn-taking).

¹⁸ Also called time coded metadata, or “strata” (Troncy, Huet, & Schenk, 2011).

¹⁹ See Sloetjes (2014) for more details about this software and Melgar et al. (in press) about its application in film studies.

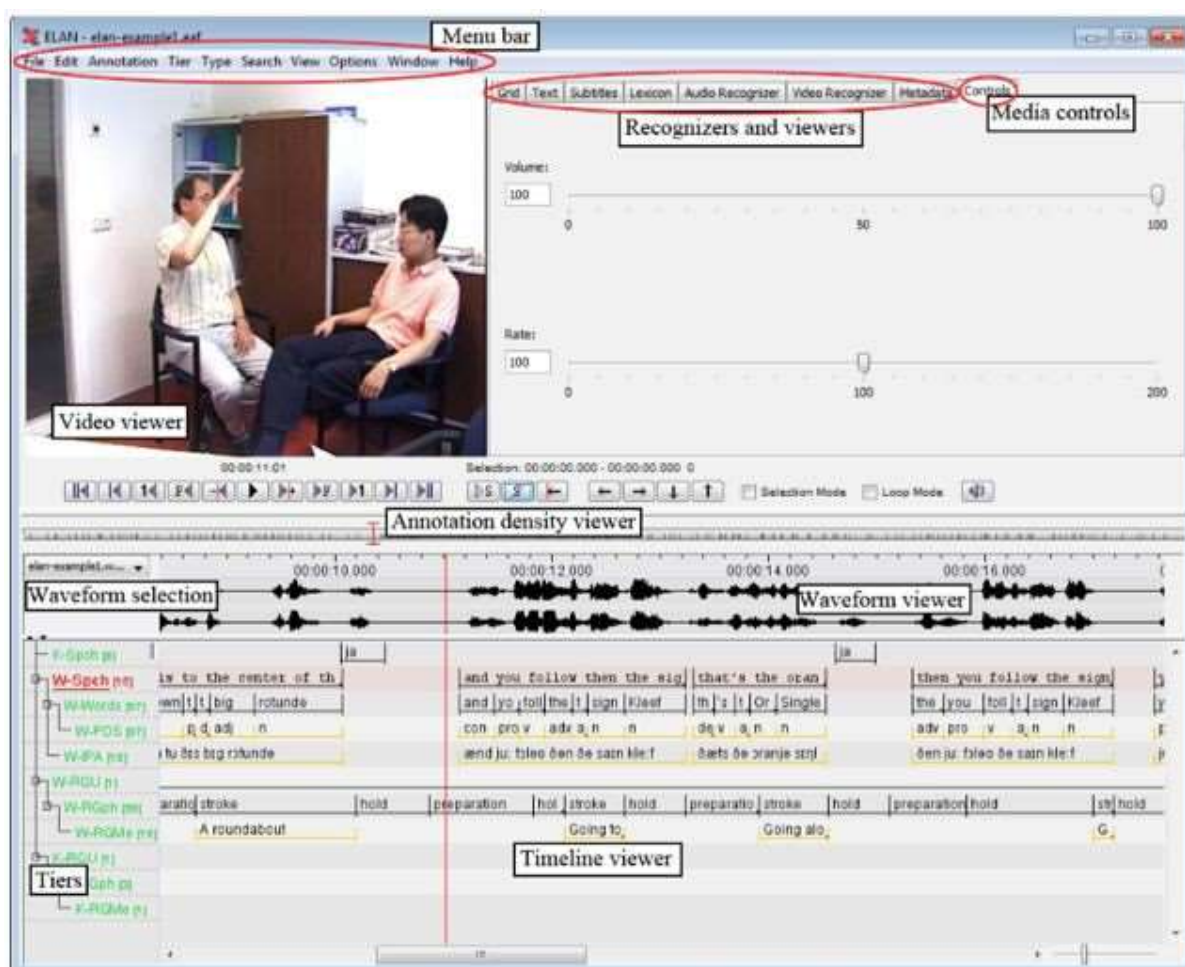


Figure 3. ELAN interface

Research Tasks and System Terminology

There is a lack of consistency and no shared terminology (or conceptualization) of describing qualitative data analysis tasks. This is reflected in the “idiosyncratic” names used for tools that are actually present in each of the software packages (see Evers, this issue). Finlayson (2016) clearly described this “lack of community-wide idiom,” explaining that this forces the users to spend more time familiarizing themselves with changing terms for the same tasks, with a negative effect in the learning process. Thus, for the purpose of comparability, the first step is to determine the correspondence between the “idiosyncratic” terms used by each package, and the concept they correspond to, either in qualitative analysis or in other information-related domains.

Table 2 shows our proposed mapping between the terminology used by the three packages we selected for review, and more widely used terms in the domain of qualitative/AV research.

Table 2. Comparison of terms used by different software packages

Object or common task	Nvivo	Transana	ELAN	Correspondence with broader qualitative/AV research domain
Objects of analysis	“Source”	“Media file” or “Episode”	“Media file”	-Resource ^(b) -Information object ^(c)
Transcribing	“Transcript”	“Transcript”	“Transcription” or “Annotation” in a wider sense	Transcript ^(d) Natural language representation ^(d)
Segmenting	“Reference” ^(e)	“Clip” ^(f)	“Time interval”, “Tier” ^(g)	Fragment ^(b)
Coding	“Node” ^(h)	“Keyword”	-“Code”: a type of “annotation”	-Code ⁽ⁱ⁾ - “Tag” ^(j) -Annotation ^(k)
Linking ^(l)	“See also links” “Hyperlinks”	“Hyperlinks”	N/A	Linking ^(m)
Commenting	“Annotation” “Memo”	“Note”	“Comment”	Commenting ^(m)

(a) As of version 8 and the Mac version.

(b) W3C (<https://www.w3.org/TR/annotation-vocab/>).

(c) (Bekiari, C., Doerr, M., Le Boef, P., & Riva, P., 2015).

(d) We use the term “transcript,” since the three packages are all using this term. However, the most appropriate concept would be “natural language representation” (as used and explained in Ingwersen, 1996), because in some cases, the textual descriptions do not intend to “transcribe” the spoken word, but to describe the audio-visual message using free written forms (for example, summarizing what a fragment is about or what can be seen in a fragment).

(e) The concept of “segment” or “fragment” is not explicitly defined in Nvivo. The term “reference” is used for the result list of all parts of a text (or all “regions” if it is a picture) that have been coded at a certain node. In the case of audio-visual media, a “transcript entry” (a timespan) could be the equivalent of a “fragment.”

(f) Also, textual fragments are called “quotes”, and, in the case of still images, “snapshots.”

(g) ELAN defines a tier as a set of annotations that share the same characteristics. DRS and ANVIL call them “Track.”

(h) Nvivo does not use the term “code” explicitly. “Coding” in NVivo is the term used to refer to “the process of gathering material by topic, theme or case. For example, selecting a paragraph about water quality and coding it at the node “water quality,” while a “node” is the container for all “references” coded with the same “node.”

(i) Saldaña, 2016.

(j) Term often used in web annotation (W3C).

- (k) A code in qualitative analysis is most often a word or short phrase that represents an attribute of the fragment with the aim of identify it or classify it, whereas annotations are often associated to “comments” (in contextualization-related tasks). In W3C (2017) terms, “annotation” refers to both “codes” and “comments” (<https://www.w3.org/TR/annotation-vocab/#annotation>).
- (l) The concept of “link” is central to contextualization tasks. However, we leave it out of this comparison, since it’s application varies greatly between software programs. See Silver and Patashnick (2011) or the ATLAS.ti manual’s description of hyper-linking.
- (m) In the W3C annotation data model, Linking and Commenting are types of “motivations.”

In the next section we use the correspondence terms to create abstract representations (data models) of how the three packages support the common tasks, as outlined in Table 2.

Comparing Software for Audio-Visual Qualitative Data Analysis Based on Research Tasks

This section introduces a comparison of AV analysis tools based on the identification of how they support the common research tasks. Figures 4, 5 and 6 show a schematic view of how the tools structure or “model” their functionalities in order to support the research tasks. This representation can be called a “data model.”^{20,21} The figures also illustrate how each software package uses different terminology for similar tasks (the “idiosyncratic” terms are in parenthesis).

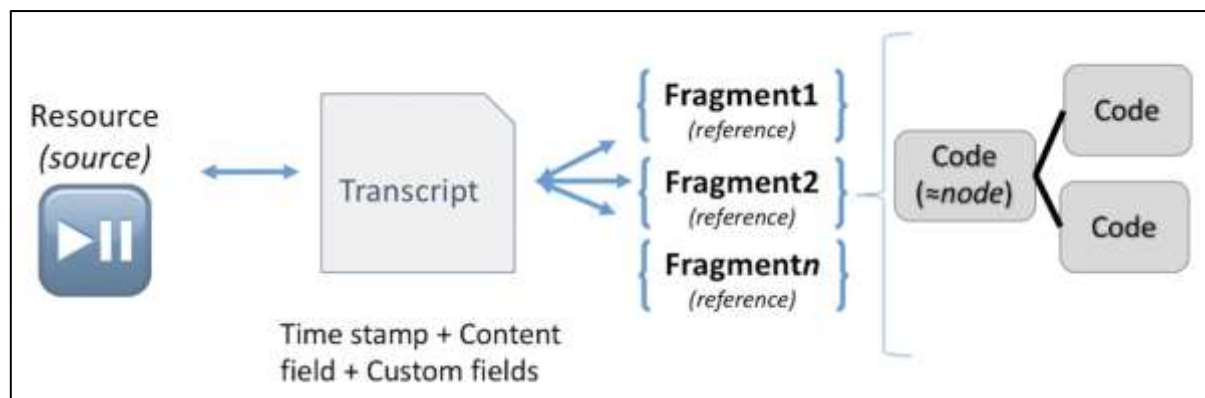


Figure 4. Abstract representation of Nvivo’s structure for supporting qualitative audio-visual analysis tasks²²

²⁰ The term “data model” refers to the design choices of each tool on implementing functionalities that support research tasks. It is a conceptual framework in which bits of data are identified and structured.

²¹ For extracting the models and documenting the functionalities, we used the latest versions of these software packages (or a demo version in the case of Transana), and their documentation as available in their websites.

²² In NVivo codes can be of two types: “theme codes,” which are concepts, and “case codes,” which contain demographic information, are classified and have attributes. In the figure, only theme codes are represented.

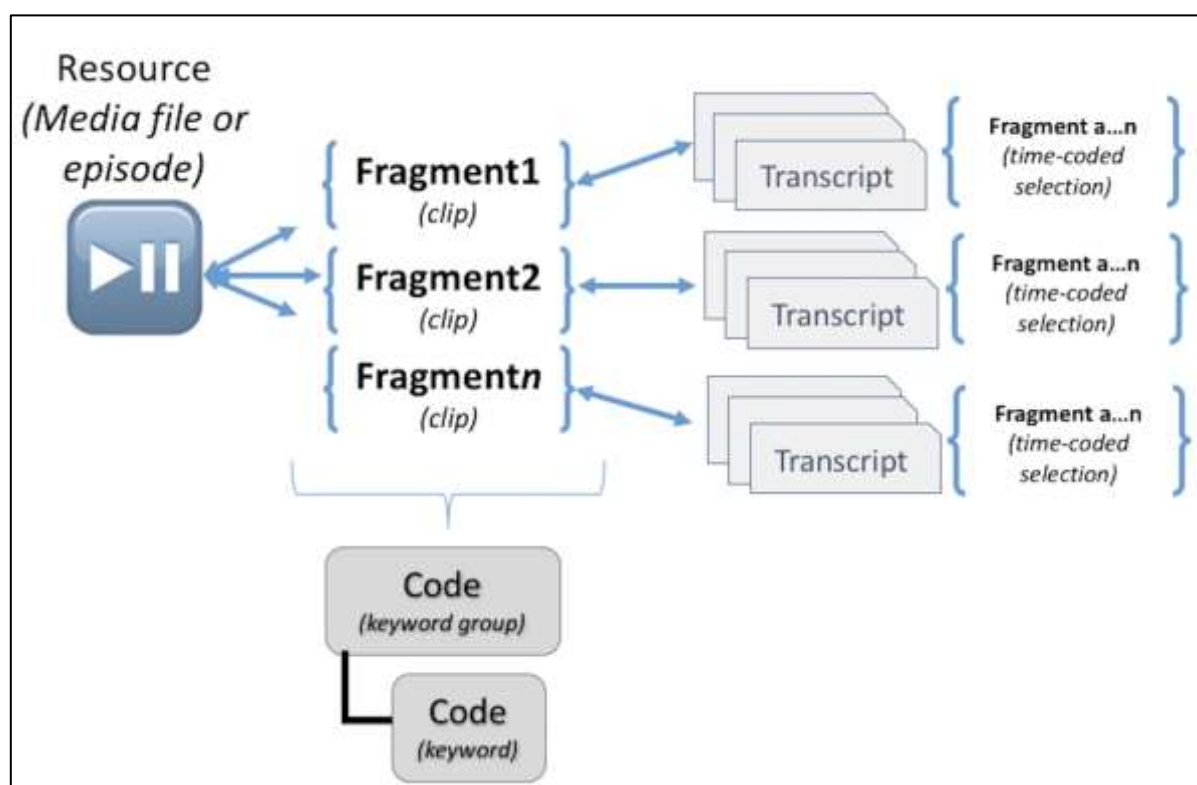


Figure 5. Abstract representation of Transana's structure for supporting qualitative audio-visual analysis tasks

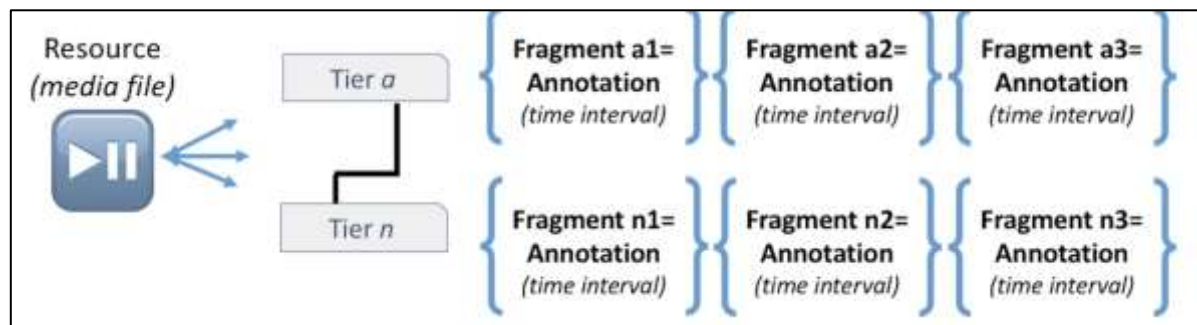


Figure 6. Abstract representation of ELAN's structure for supporting qualitative audio-visual analysis tasks

Concrete Functionalities That Support Research Tasks

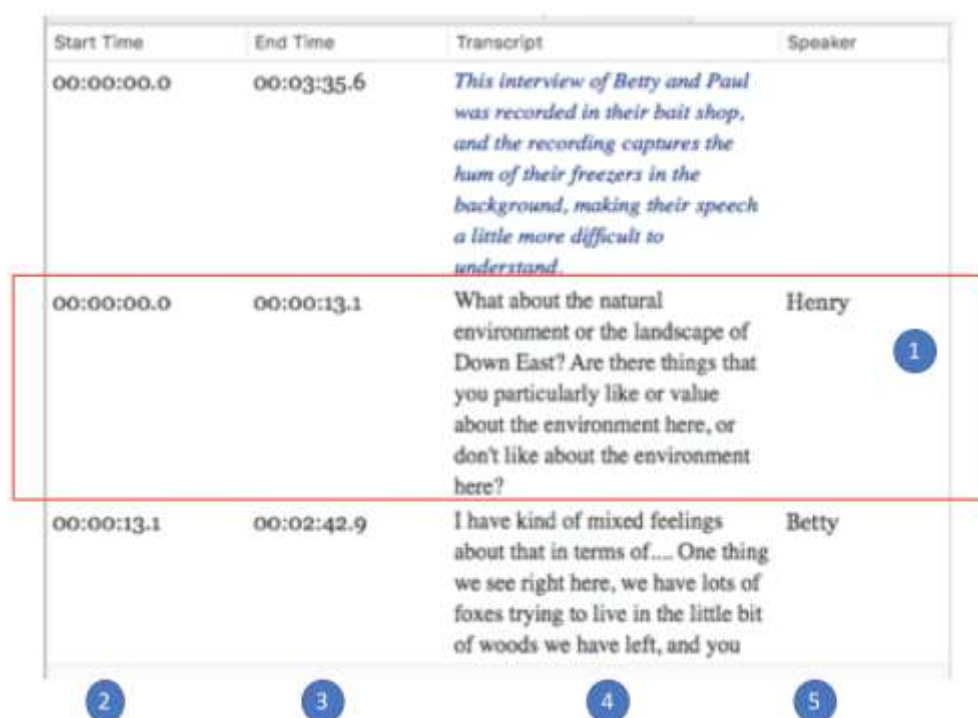
This section explains the models presented above, and how these models influence the analysis tasks. We use the research case of Alice, who is an oral historian with a memory studies focus.²³ By recording video interviews with 20 Indonesian war veterans in The Netherlands, she investigated how people construct memories of historical events. In addition, she collected documents and digitized archival material.

²³ This is an emergent field of study, see for instance: <http://www.memorystudiesassociation.org/>

Transcribing or Creating Natural Language Representations²⁴

Alice starts with importing the AV materials, which can be done in any of the three tools. She then considers whether to annotate them directly (via the video timeline) or to use a transcript as an intermediary document. She discovers that automatic speech recognition services (i.e., converting the audio speech into text) are not offered by any of the three tools (except from some recognizers included in ELAN), and that if she prefers to analyze the spoken words via a transcript, she will have to produce a verbatim transcript herself or by using an external service. She could also use any of the tools to create a transcript herself, including summarizations, or import transcripts from an external service into the software.

She observes that in **NVivo**, only one transcript can be synchronized per media file. Other transcripts can be imported into the project file, but only as stand-alone, non-associated documents. She notices that, in **NVivo**, synchronized transcripts are structured as a table in which each row is a “transcript entry” (Figure 7, 1). Transcript entries have a minimum of four columns: 2) Start time, 3) End time, 4) Transcript text or “content”, and 5) “Speaker” column, or “Custom” field. Thus, she will have to choose whether to use the “content” column of NVivo’s transcript to enter, for example, a verbatim transcript, a Jeffersonian transcript, or a gisted transcript.



Start Time	End Time	Transcript	Speaker
00:00:00.0	00:03:35.6	<i>This interview of Betty and Paul was recorded in their bait shop, and the recording captures the hum of their freezers in the background, making their speech a little more difficult to understand.</i>	
00:00:00.0	00:00:13.1	What about the natural environment or the landscape of Down East? Are there things that you particularly like or value about the environment here, or don't like about the environment here?	Henry
00:00:13.1	00:02:42.9	I have kind of mixed feelings about that in terms of.... One thing we see right here, we have lots of foxes trying to live in the little bit of woods we have left, and you	Betty

2 3 4 5

Figure 7. Elements of a transcript in NVivo

Transana, in contrast, allows multiple transcripts (i.e., up to five documents or data files) to be associated with the same media object (or fragment), requiring at least one transcript to be associated with each audiovisual data file. In Figure 2, each of the three horizontal panels in the bottom left is a transcript file, which corresponds to a different type of transcript. In **ELAN**, each tier could be considered as a type of transcript (in its broader sense of “natural language representation”), since it is possible to add longer textual annotations (not just codes) to each fragment, thus, this would allow her to create multiple (unlimited) transcripts, all connected to the media file. Also, as shown in Figure 8, each part of a transcript (e.g., an utterance) can have

²⁴ See note (d) under Table 2 for an explanation of this concept.

an additional transcript type (e.g., a phonetic transcription of the utterance) representing an annotation on top of another annotation.

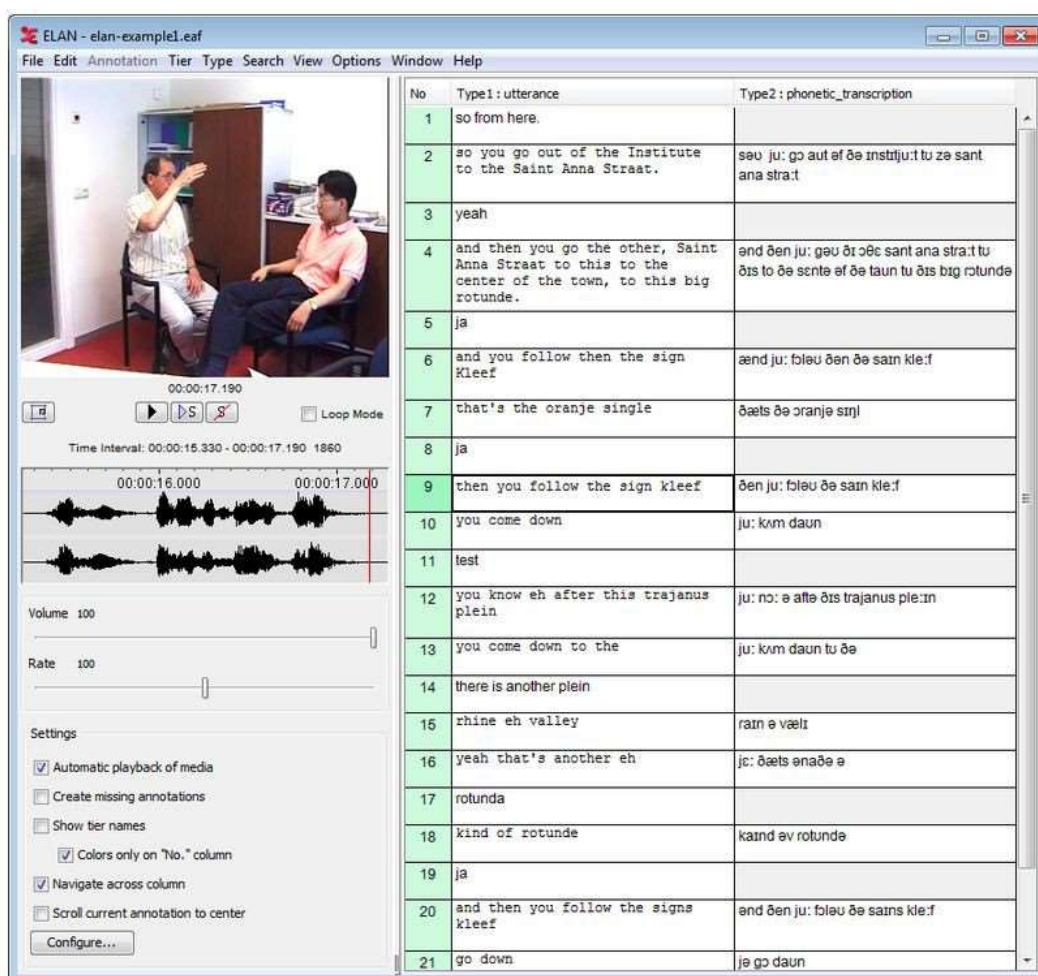


Figure 8. ELAN's way to provide support for multiple transcripts. The left column contains a textual annotation of a fragment, which has a corresponding "phonetic transcription" on the right column

In sum, when selecting a tool, Alice concludes that attention must be paid to the type and number of transcripts that the tool can handle (in relation to what she needs for her project), as well as the available synchronicity between transcript and AV file, and between the transcripts when there is more than one.²⁵

Segmenting

Alice would like to segment each interview as part of her analysis. For each segment, she would like to separate the analytical elements, which in her project are: the actual spoken words, and her annotations related to gestures. The three software packages offer different ways of segmenting a media object: NVivo's "tabular presentation" (Silver & Patashnick, 2011, p. 14; and Figure 7) converts every "transcript entry" (a time span) into a fragment, and each change in a timespan is treated as a new fragment. As a result, segmenting in NVivo has limited

²⁵ See also Silver and Patashnick (2011) for more details about how synchronicity supports research tasks.

functionality for multi-level segmentation, and that establishing connections (e.g., hierarchies) between the segments is not straightforward.

She observes that in **Transana** it is possible to have a fragment (i.e., a “clip”) with several time spans or time-coded information within it, which provides more flexibility in terms of granularity. Figure 9 shows a verbatim transcript on the left side and a gesture “transcript” on the right side, both with different time anchors.

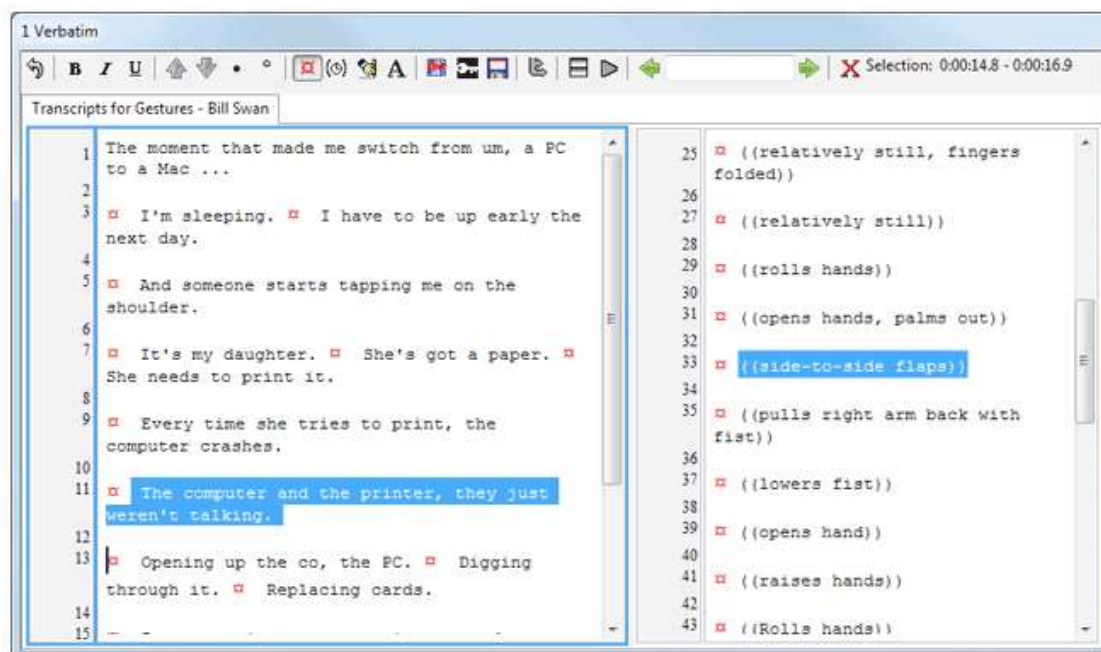


Figure 9. Transana's multiple transcript synchronization

Finally, Alice notices that **ELAN** offers multi-level segmentation through its tiered approach. That is, fragments are created horizontally, along the timeline, and vertically, across different tiers. Also, **ELAN** offers some automatic options to analyze the audio signal (e.g., to identify speaker turns or silences), which could assist with creating the segments (units of analysis), saving her some time. Neither NVivo nor Transana provide this option, even though some elements could be achieved via workarounds, for instance, by using the “speaker” column in NVivo to indicate the name of a “tier”. However, the visualization of these dependencies in the coding stripes are limited (see Figure 1, 4 and 5 panels with coding stripes).

Thus, Alice concludes, NVivo would not allow her to create broader segments for topics which would also include smaller dependent segments with gestures. If using Transana, she could achieve this multi-level segmentation by using two transcripts segmented accordingly (one for topics, and one for gestures). ELAN offers the most elaborate segmenting possibilities. There are, however, important differences at the level of coding and contextualizing which will also impact her decision.

Coding

Alice observes that codes in NVivo can be assigned directly via the timeline or the transcript representation (“content” column), by selecting parts of text within a transcript. The process of refining, aggregating, editing, or reordering codes is supported in a very flexible way, since the codes are listed separately as entries in an index, while keeping their original locator information (Figure 3, panel 3). She notices that she could create levels and groups of codes to form a hierarchy, and that the values entered in the “speaker” column could be transformed into codes, which she could use,

for instance, to indicate a category for each fragment (e.g., gesture fragment, spoken word fragment). The hierarchies, and this “speaker” column could partially help to overcome NVivo’s lack of explicit support for multi-level segmentation.

She also considered creating different transcripts in Transana to emulate the tiered approach, but she discovers that the flexibility of thematic analyses and bottom-up annotation is limited in this program since it only offers a two-level hierarchy of codes (“keywords”). Also, her multi-level segmentations (using the different transcripts) can be displayed only in a roundabout way, through the “keyword sequence map” offered as an output to visualize sequences of codes according to keyword groups (Figure 10), as well as through the “hybrid” visualization of keywords in the timeline.

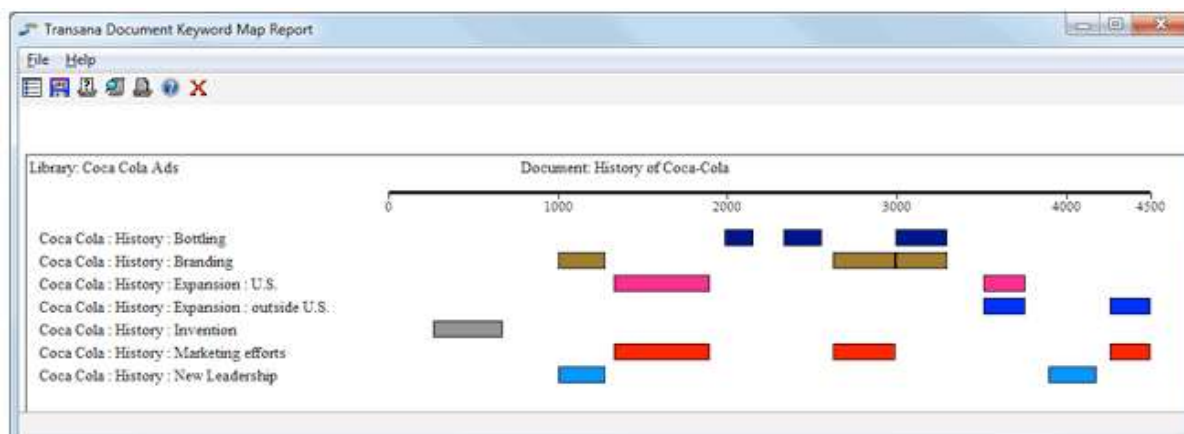


Figure 10. “Document keyword map” in Transana

Turning to ELAN Alice realizes that, in contrast to NVivo and Transana, coding is always done directly (via the timeline) and not via a transcript (Figure 11). This allows the researcher to enter any type of annotation (a code, or a broader piece of text or transcript).

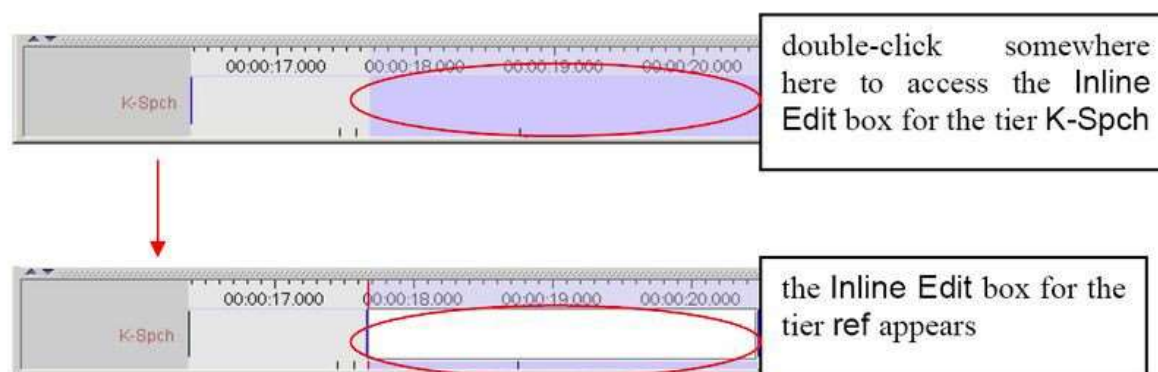


Figure 11. Inline edit box for the tier in ELAN

Even though these tiers with annotations can be clearly visualized straight away, a downside is that creating bottom-up coding of the themes or topics is not flexible enough, since she would have to either use a pre-defined controlled vocabulary, which she does not have in advance, or be constantly editing the annotations as they emerge while listening and viewing the interviews. This will be difficult if Alice needs to constantly rearrange annotations from the same tier into hierarchical analytic categories.

Linking

Besides her segmenting and coding tasks, Alice notes that her contextualization-related tasks are significantly influenced by each tool. For instance, in NVivo she could easily import her photos, newspaper articles, and letters, which she could link internally in different ways, e.g., by identifying topics (themes) using the same set of emergent themes from the interviews, or externally, via hyperlinks. This could help her contextualize the specific events that an interviewee was referring to (for instance by connecting a vague mention to a battle in a village with a complete newspaper article where this battle is reported). This could also be done in Transana, through simultaneous analysis of textual data and still images together with AV media by using the same keywords, using only a two-level hierarchy of codes. Finally, since ELAN is devoted to analyzing audio or video files, associating AV media files to related textual or visual objects is not possible.

Commenting

Finally, in all three software packages Alice can record analytical insights gained throughout the project in the form of the idiosyncratically named “memos,” “annotations” (in a narrow sense), “notes,” or “comments.”

Methodological Implications

After this review Alice concludes that she will have to decide how much emphasis to give to formal aspects of the data (i.e., in identifying the gestures very precisely), or to the themes, and how important the use of simultaneous annotation of the different media gathered will be for her investigation. AV-centered scholars often consider the media they analyze as “texts,”²⁶ examining their stylistic features, themes or narrative elements, and interpreting how “made meanings” or representations are structured (Rose, 2016). Scholars who make intensive use of AV media (e.g., visual anthropologists, oral historians) do so to understand, for instance, aspects of behavior and culture.

Two factors in particular influence research using qualitative analysis software. First, is the selection of a unit of analysis. In this sense, the segmenting task will be influenced by the perspectives of different academic traditions (for example, the use of shots as formal unit of analysis in film studies, or of words or morphemes for the study of language). Even though the three analyzed software packages are agnostic in relation to which unit of analysis is chosen to segment (i.e., one can choose any start and end point to make the fragments), the way to structure, relate, and enable annotation of these fragments differs. ELAN, for example, is specifically designed with a tiered approach, in which segments and their annotations belong to a specific tier or facet. Users of NVivo and Transana may find workarounds to simulate this approach, however visualization of these data via the timeline does not properly support it. This is important because, in certain cases, AV-centered scholars focus on one dimension (modality) of the audiovisual message only, for instance, on spoken words via the analysis of the audio signal. Media scholars may analyze discourse or coverage of specific events, debates, or groups of people, and oral historians may focus on the words used when people narrate historical events. However, regardless of the emphasis, most scholars who use AV media assume a “multimodality” perspective (Schmidt et al., 2009), looking at co-occurrences in different dimensions, for instance, when analyzing non-verbal behavior (e.g., hand movements vs. facial expression), or analyzing recurring characters or motifs in relation to stylistic aspects (e.g., city buildings vs. camera movements), or spoken words in relation to gestures (Alice’s example). Thus, when using software

²⁶ See Kirkegaard (2008), Noordegraaf (2016), and Melgar et al. (2017).

tools for qualitative analysis, it is essential to pay attention to what extent the software supports the selection of analytical units not only for thematic analysis but also for these multi-modal analyses.

Second, contextualization-related tasks (i.e., linking and commenting) are key to qualitative research. In this regard, QDAS packages give better support than the professional audiovisual annotation tools, since they enable qualitative analyses of most types of media within the same application. While it is known that contextualization in media studies is an essential scholarly task (Bron et al., 2015), more studies need to be done to understand whether it is essential for scholars using a professional video analysis tool for detailed AV analysis to also analyze accompanying texts or still images with the same tool, or whether it is more convenient to use specialized software for each media type and perform cross-media analyses with the resulting annotations generated by each software.

None of the three packages reviewed in this paper fully support these two methodological requirements for qualitative AV-analysis. Thus, performing certain tasks with one specific package at a time would be a way to overcome the limitations of software packages and to take advantage of their strengths.²⁷ However, this requires the ability to exchange data between applications,²⁸ and the ability of scholars to develop the expertise to work with data processing at a more general level, rather than be tied to a specific tool.

Conclusion

Our paper aimed to demonstrate the impact of software tools and their underlying design choices on the workflow and outcome of AV data analysis. Qualitative researchers need to carefully weigh their options based on their research goals and methodologies. In particular, AV-centered scholars can to some extent control details of the tool-based research workflow by using multiple systems, switching between tools after certain pre-processing and/or analysis steps and selecting the optimal software for subsequent steps. This requires not only a good understanding of the impact of tool design in the research process, but also knowing which tools are available, how they can complement each other and to what extent data can be exchanged between them. Since the design and interoperability of tools affect qualitative research methodology, what on the surface may seem just a technical challenge requires more in-depth methodological discussions and engagement by researchers, shifting from passive users to active participants in the conceptualization and modeling of their research instruments.

References

- Bekiari, C., Doerr, M., Le Boef, P., & Riva, P. (Eds.). (2015). *FRBR: Object-oriented definition and mapping from FRBRER, FRAD and FRSAD* (Version 2.2). International Working Group on FRBR and CIDOC CRM Harmonisation. Retrieved from https://www.ifla.org/files/assets/cataloguing/frbr/frbroo_v2.2.pdf
- Benardou, A., Dallas, C., Papaki, E., Constantopoulos, P., Angelis, S., Baldwin, T., ... Sjögren, B. (2013). *Deliverable D1.2 – State of the art report on digital research practices, tools and scholarly content use*. The Hague, Netherlands: Europeana Professional.

²⁷ For the KWALON 2016 conference discussion about the need for multiple tools, see <https://youtu.be/sU2hv4N6d6I?t=2543>, about a desire for a common core of functionalities, see <https://youtu.be/sU2hv4N6d6I?t=1105>.

²⁸ This is what is commonly referred as to the “interoperability” problem (van de Sompel, 2015), which has been around since the tools were created. For interoperability issues among QDAS, see, for instance, Evers (this issue), Schmidt (2009), Evers et al. (2011), Finlayson (2016) and, Rizkallah (2016).

- Berry, D. M. (2012). Introduction: understanding the digital humanities. In D. M. Berry (Ed.), *Understanding Digital Humanities* (pp. 1–20). Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan. Retrieved from [DOI: 10.1057/9780230371934](https://doi.org/10.1057/9780230371934)
- Borek, L., Dombrowski, Q., Perkins, J., & Schöch, C. (2016). TaDiRAH: A case study in pragmatic classification. *Digital Humanities Quarterly*, 10(1). Retrieved from <http://www.digitalhumanities.org/dhq/vol/10/1/000235/000235.html>
- Bron, M., van Gorp, J., & de Rijke, M. (2015). Media studies research in the data-driven age: How research questions evolve. *Journal of the Association for Information Science and Technology*, 67(7), 1535-1554. doi:10.1002/asi.23458
- Chu, C. M. (1999). Literary critics at work and their information needs: A research-phases model. *Library & Information Science Research*, 21(2), 247–273. [https://doi.org/10.1016/S0740-8188\(99\)00002-X](https://doi.org/10.1016/S0740-8188(99)00002-X)
- Clivaz, C. (2016). Images, sound, writing in Western: A long hatred-love story? Keynote address given at *Audiovisual Data and Digital Scholarship: Towards Multimodal Literacy*. Krakow, Poland: SIG.
- Dasiopoulou, S., Giannakidou, E., Litos, G., Malasioti, P., & Kompatsiaris, Y. (2011). A survey of semantic image and video annotation tools. In G. Paliouras, C. D. Spyropoulos, & G. Tsatsaronis (Eds.), *Knowledge-driven multimedia information extraction and ontology evolution* (Vol. 6050, pp. 196-239). Berlin, Germany: Springer.
- Desrochers, N., & Apollon, D. (Eds.). (2014). *Examining paratextual theory and its applications in digital culture*. United States: IGI Global.
- Evers, J. C. (2011). From the past into the future. How technological developments change our ways of data collection, transcription and analysis. *Forum Qualitative Sozialforschung*, 12(1). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1636>
- Evers, J. C. (2015). Elaborating on thick analysis: About thoroughness and creativity in qualitative analysis. *Forum Qualitative Sozialforschung*, 17(1). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/2369>
- Evers, J. C., Silver, C., Mruck, K., & Peeters, B. (2011). Introduction to the KWALON experiment: Discussions on qualitative data analysis software by developers and users. *Forum Qualitative Sozialforschung*, 12(1). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1637>
- Finlayson, M. A. (2016). *Report on the 2015 NSF workshop on unified annotation tooling*. Miami, FL: Florida International University & Massachusetts Institute of Technology.
- Genette, G. (1997). *Palimpsests: Literature in the second degree*. Lincoln, NE: University of Nebraska Press.
- Gibbs, G. R. (2014). Using software in qualitative analysis. In U. Flick (Ed.), *The Sage handbook of qualitative data analysis* (pp. 277-294). London, UK: Sage.
- Gibson, W., Callery, P., Campbell, M., Hall, A., & Richards, D. (2005). The digital revolution in qualitative research: Working with digital audio data through Atlas.Ti. *Sociological Research Online*, 10(1). doi: 10.5153/sro.1044
- Giunti, L. (2014). L'analyse du film à l'ère numérique. Annotation, geste analytique et lecture active [The analysis of film in the digital age: Annotation, analytical gesture and active reading]. *Cinéma & Cie*, XIV(22/23), 127-144.
- Gey, R. (Ed.). (n.d.). Social Science Software: Software in Social Science Research (SoSciSo). Retrieved from <http://www.sosciso.de/en/>
- Hewett, R., & Barber, S. (Eds.). (2013). *Audiovisual citation: BUFVC guidelines for referencing moving image and sound*. British Universities Film & Video Council. Retrieved from <http://bufvc.ac.uk/wp-content/media/2018/01/Learning-on-Screen-AV-Citation-2017-ONLINE.pdf>

- Huurnink, B., Snoek, C., de Rijke, M., & Smeulders, A. (2012). Content-based analysis improves audiovisual archive retrieval. *IEEE Transactions on Multimedia*, 14(4), 1166-1178. doi: 10.1109/TMM.2012.2193561
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 3-50. doi: 10.1108/eb026960
- Ingwersen, P., & Järvelin, K. (2005). *The turn: Integration of information seeking and retrieval in context*. Dordrecht, The Netherlands: Springer.
- Kendall, D. E. (2012). *Sociology in our times: the essentials* (8th ed.). Belmont, CA: Thomson/Wadsworth.
- Kirkegaard, B., & Borlund, P. (2008). Characteristics of information needs for television broadcasts of scholars and students in media studies. In P. Borlund (Ed.), *Proceedings of the second international symposium on Information interaction in context* (pp. 116-122). New York, NY: ACM Press.
- Lejeune, C. (2013). A Brief History of Software Resources for Qualitative Analysis. In B. Reber & C. Brossaud (Eds.), *Digital Cognitive Technologies* (pp. 169–186). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118599761.ch11>
- Marsden, A., Mackenzie, A., Lindsay, A., Nock, H., Coleman, J., & Kochanski, G. (2007). Tools for searching, annotation and analysis of speech, music, film and video – A survey. *Literary and Linguistic Computing*, 22(4), 469-488. doi:10.1093/lc/fqm021
- Melgar Estrada, L., Hielscher, E., Koolen, M., Olesen, C., Noordegraaf, J., & Blom, J. (in press). Film analysis as annotation: Exploring current tools and their affordances. *The Moving Image: The Journal of the Association of Moving Image Archivists*.
- Melgar Estrada, L., Koolen, M., Huurdeman, H., & Blom, J. (2017). *A process model of time-based media annotation in a scholarly context*. Presented at the CHIIR 2017: ACM SIGIR Conference on Human Information Interaction and Retrieval, Oslo, Norway.
- Nixon, L., & Troncy, R. (2014). Survey of semantic media annotation tools for the web: Towards new media applications with linked media. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, & A. Tordai (Eds.), *The semantic web: ESWC 2014 satellite events* (Vol. 8798, pp. 100-114). Cham, Switzerland: Springer.
- Noordegraaf, J. (2016). Computational research in media studies: Methodological implications. *KWALON*, 21(1), 52-59.
- Palmer, C. L., Tefteau, L. C., & Pirmann, C. M. (2009). *Scholarly information practices in the online environment: Themes from the literature and implications for library service development*. Dublin, Ohio: OCLC Research.
- Paulus, T., & Lester, J. (2016). ATLAS.ti for conversation and discourse analysis. *International Journal of Social Research Methodology*, 19(4), 405-428.
- Pickard, A. J., & Childs, S. (2013). Grounded theory: Method or analysis? In D. Giddens (Ed.), *Research methods in information* (pp. 177-188). Chicago, IL: Neal-Schuman.
- Rizkallah, E. (2016). *QDA software compatibility: Towards an exchange format with developers for their users*. Presented at the Kwalon: Reflecting on the future of QDA Software, Rotterdam, The Netherlands.
- Rose, G. (2016). *Visual methodologies: An introduction to researching with visual materials*. Thousand Oaks, CA: Sage.
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Thousand Oaks, CA: Sage.
- Sandom, C., & Enser, P. G. B. (2001). VIRAMI: *Visual information retrieval for archival moving imagery*. Presented at the International Cultural Heritage Informatics Meeting, Archives & Museum Informatics, Milano, Italy.
- Schmidt, T., Duncan, S., Ehmer, O., Hoyt, J., Kipp, M., Loehr, D., ... Sloetjes, H. (2009). An exchange format for multimodal annotations. In M. Kipp, J. C. Martin, P. Paggio, & D.

- Heylen (Eds.), *Multimodal corpora* (Vol. 5509, pp. 207-221). Berlin, Germany: Springer.
- Schnettler, B., & Raab, J. (2008). Interpretative visual analysis. Developments, state of the art and pending problems. *Forum Qualitative Sozialforschung*, 9(3). Retrieved from <http://dx.doi.org/10.17169/fqs-9.3.1149>
- Silver, C., Bulloch, S. L., & Rivers, C. (2011). *Comparative techniques for analysing audio-visual data in selected CAQDAS packages*. Surrey, UK: University of Surrey.
- Silver, C., & Lewins, A. (2014). *Using software in qualitative research: A step-by-step guide* (2nd ed.). Los Angeles: Sage.
- Silver, C., & Patashnick, J. (2011). Finding fidelity: Advancing audiovisual analysis using software. *Forum Qualitative Sozialforschung*, 12(1). <http://dx.doi.org/10.17169/fqs-9.3.1149>
- Sloetjes, H. (2014). ELAN: Multimedia annotation application. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 305-320). Oxford, UK: Oxford University Press.
- Stanczak, G. C. (Ed.). (2007). *Visual research methods: Image, society, and representation*. Thousand Oaks, CA: Sage.
- Tesch, R. (1990). *Qualitative research: Analysis types and software tools*. Hoboken, NJ: Taylor and Francis.
- Troncy, R., Huet, B., & Schenk, S. (2011). *Multimedia semantics: Metadata, analysis and interaction*. Chichester, UK: John Wiley & Sons.
- UNESCO. (2012). *Documentary and audiovisual heritage*. Retrieved from www.unesco.org/new/en/phnompenh/communication-and-information/documentary-and-audiovisual-heritage/
- Usai, P. C., Francis, D., Horwath, A., & Loebenstein, M. (Eds.). (2008). *Film curatorship: Museums, curatorship and the moving image*. Vienna, Austria: Austrian Film Museum.
- Unsworth, J. (2000). *Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this?* Presented at the Symposium on Humanities Computing: Formal Methods, Experimental Practice, London, UK, King's College.
- van de Sompel, H., & Nelson, M. L. (2015). Reminiscing about 15 years of interoperability efforts. *D-Lib Magazine*, 21(11/12). doi:10.1045/november2015-vandesompel
- W3C. (2017, February 23). *Web annotation data model: W3C Recommendation*. Retrieved from <http://www.w3.org/TR/annotation-model/>
- Weigel, C. (2016). *Report on state of the art for moving image analysis* (Innovative e-environments for Research on Cities and the Media No. D7.1, v1.3).
- Woods, M., Paulus, T., Atkins, D. P., & Macklin, R. (2016). Advancing qualitative research using qualitative data analysis software (QDAS)? Reviewing potential versus practice in published studies using ATLAS.ti and NVivo, 1994-2013. *Social Science Computer Review*, 34(5), 597-617. doi:10.1177/0894439315596311

Tools

- ELAN for Mac OS 4.9.4. [Computer software]. (2016). Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics, The Language Archive.
- NVivo for Mac 11.3.2 (starter edition). [Computer software]. (2016). Doncaster, Victoria: QSR International.
- Transana 3.01 [Computer software]. (2016). Madison, WI: Spurgeon Woods LLC.

Author Note

Liliana Melgar holds a Ph.D. in Information Science and works as a postdoctoral researcher at the University of Amsterdam, where she conducts the user studies for the CLARIAH project, the Dutch digital humanities infrastructure. Research interests include information behaviour, scholarly annotations, multimedia access and retrieval, classification theory, and media studies. Correspondence regarding this article can be addressed directly to: melgar@uva.nl

Marijn Koolen is a scientific software engineer at Huygens ING, Royal Netherlands Academy of Arts and Sciences (The Netherlands), with a PhD in Information Retrieval. Research interests include information behaviour, social annotations, multimedia access and retrieval, digital humanities pedagogy, and digital literary studies. Correspondence regarding this article can also be addressed directly to: marijn.koolen@huygens.knaw.nl

We would like to thank our reviewers for their valuable input.

Copyright 2018: Liliana Melgar-Estrada, Marjin Koolen, and Nova Southeastern University.

Article Citation

Melgar-Estrada, L., & Koolen, M. (2018). Audiovisual media annotation using qualitative data analysis: A comparative analysis. *The Qualitative Report*, 23(13), 40-60. Retrieved from <http://nsuworks.nova.edu/tqr/vol23/iss13/4>
