Susan Hogervorst

CrossEWT: *Cross-Medial Analysis of World War II Eyewitness Testimonies*

Abstract

The CLARIAH research pilot project ‘CrossEWT’ entailed an inventory of interviews with eyewitnesses of the Second World War as published in newspapers, television documentaries and oral history databases. The aim was to analyse the topics of these interviews, and how these may have differed per medium and over time since 1945. The Media Suite facilitates comparative research into collections of different media types that traditionally are studied separately. Especially for oral history research, this is innovative and promising, in at least two ways. First, it potentially stimulates the re-use of interviews, not only by media scholars who might accidentally discover this material, but by (oral) historians as well. Second, a video annotation tool is provided that enables analysing the visual, non-verbal and performative aspects of oral history interviews, while traditional oral history research often heavily relies on the transcript.

Keywords

oral history, video annotation, digital humanities, World War II memories

Since the 1980s, survivors and eyewitnesses of the Second World War, and especially the Holocaust, have become increasingly prominent in historical culture throughout the western world. This advent of the witness, as the French historian Annette Wieviorka framed it,¹ was set in motion by one of the first global media events on television: the trial against Adolf Eichmann in Jerusalem, 1961.² The live broadcast of the many compelling testimonies of the witnesses, and the accompanying images of the camps that flooded the media, commanded world attention, and put the Holocaust onto the public record.³ Anticipating their forthcoming disappearance, many initiatives have been undertaken to preserve eyewitnesses’ memories for the future: in documentary films, oral history archives, educational projects, memoirs, newspaper articles, talk shows, books, and exhibitions. However, there is no systematic research about which topics have actually been addressed in their accounts. Such knowledge

could not only enhance our understanding of the different socio-cultural functions and values attributed to testimonies. It could also help disclosing the many recorded testimonies in what some call the post-witness era.⁴

CrossEWT entails a cross-media and diachronic content analysis of eyewitness testimonies (EWT's) about World War II. The focus is on testimonies that have been produced in the Netherlands since 1945, in three different media constellations: newspapers, television programmes and oral history interviews. The CLARIAH Media Suite brings together collections of these different types and therewith allows for such a comparative investigation. The aim of the project was to investigate which topics these testimonies contain; how this might differ per medium/genre, and how this may have changed over time.

Oral history, the digital revolution, and interdisciplinarity

While there is an increasing amount of work based on archived oral history interviews, and although the issue of re-use has been discussed from different perspectives by scholars from a variety of disciplines,⁵ most research within the field is based on interviews conducted by the researcher her- or himself. Moreover, oral history as a field seems to operate rather separately from other branches of the historical discipline, as well as from other disciplines that use interviews as qualitative method.⁶ In this sense, the fact that the Media Suite brings together the 'regular' audiovisual collections from the Netherlands Institute of Sound and Vision, and video interviews from various Dutch oral history collections, is a major contribution.

As Scagliola and De Jong have argued, the digital revolution has shifted the balance between spoken and written text.⁷ Digital interview transcripts, sometimes automatically generated, allow for text mining oral history interviews, and comparing them with other textual sources.⁸ Nonetheless, although many oral history interview collections have been made accessible and even searchable online, their potential for academic research (let alone non-academic dissemination) is not yet fully exploited, and large scale comparative analysis of oral history interviews remains challenging. This was also apparent in CrossEWT.

The main challenge was the creation of the corpus, for which the Media Suite was to be used. The issues in corpus creation will therefore be the focus of the remainder of this chapter. For the analysis, the corpus had to be exported to external tools – i.e. outside of the Media Suite – in order to perform some simple text mining tasks, such as determining word frequencies and collocations. In this way, a first systematic overview of the content of eyewitness accounts varying over time and per genre could be generated. One of the outcomes of this analysis was that, regardless of the medium, interviews with witnesses about World War II quite often seem to address family relations, especially in more recent decades. This would be in line with the increased public attention to individuals and personal emotions in the media in general, and with the perception of World War II eyewitness testimonies as embodying authentic emotions in particular.⁹

Corpus creation: Oral history interview transcripts and the problem of metadata

The corpus for CrossEWT consisted of three sub-corpora: one set of transcripts of oral history interviews with eyewitnesses of World War II; one set of Dutch newspaper articles that have explicitly been derived from interviews with such eyewitnesses, published since May 1945; and one set of transcripts of all documentaries about the war that have been broadcast on television in the Netherlands and that contain one or more interviews with a World War II witness. This section will discuss the issues in creating the oral history sub-corpus.

At the outset of this pilot project, in the summer of 2017, 49 oral history collections (sometimes called ‘datasets’) were available in the Media Suite. The interview collections themselves have been deposited at DANS (Data Archiving and Networking Services), and are listed in the CKAN registration system – a general, internationally used data management software application, often used for public archives. When entering the Media Suite, and looking for oral history data, one selects the tag ‘oral history’ in CKAN, which at the time gave 49 results. 21 results (‘datasets’ were retrieved when ‘oorlog’ was used as a query:

Figure 1. Screenshot of search results for 'oorlog' in oral history collection in CKAN, (Collection inspector, Media Suite version 3, July 2018).

Besides the fact that this particular way of presenting the results does not make clear how many interviews these retrieved datasets actually contain, it is far from obvious to trace specific interviews or subjects in the metadata in CKAN. These are divided into multiple layers, such as collection names, 'groups', and datasets. These collection names are often rather poetic or catchy, such as 'Bommen en habijten' ('bombs and habits') or 'Reis van de Razzia' ('journey of the raid'), or, contrarily, rather brief and restrained, such as 'Collectie Diederichs' ('collection Diederichs'). In most cases, these were the titles of the originating interview projects, that had to be somewhat appealing in the context of the necessary funding applications. In other cases, the collections were named after their depositor, often the researcher or journalist who had conducted the interviews. Either way, these collection titles do not always fully reflect the (range of) the content of the interviews in a way that is clear for researchers. The original interview titles remain important, of course, to be able to trace back an interview to its original context, but when searching for specific items, collection titles such as 'Heropvoeding jeugdige politieke delinquenten' ('re-education of young political offenders') or 'Verteld verleden' ('narrated past') complicate rather than facilitate the search process.

Figure 2. Screenshot results oral history collection, query ‘Tweede Wereldoorlog’, Single Search tool (Media Suite version 3, July 2018).

The next step, after having acquired some idea of relevant collections, is to search the actual content of the interviews (and the item-level metadata). This is done via the Single Search tool in the Media Suite. In this environment, collections can be searched by search words (with or without Boolean operators), and retrieved items can be accessed. When searching for ‘Tweede Wereldoorlog’ (‘World War II’) in the oral history collections, we get 1105 results:

This means that the 21 datasets we found in CKAN, in total seem to contain no less than 1105 individual interviews. The results are sorted by creator, media type, thematic collection and subject. Remarkable, and confusing, is the fact that, of these 1105 retrieved interviews, only 655 interviews are labelled as ‘World War II’. Moreover, since these categories and their results are listed in their entirety, it is a very long list to scroll down. More importantly, these results have the same clarity issues as outlined above: some overlap, and some occur more than once (e.g. ‘veteraneninstituut’ (‘veterans institute’) and IPNV – Interview Portal Nederlandse Veteranen (‘interview portal Dutch veterans’) refer to the same interviews), and some simply seem irrelevant (e.g. labelled ‘oral history’). Above all, some metadata fields are fully irrelevant, such as

Figure 3. Screenshot part of results list oral history collection, query 'Tweede Wereldoorlog', Single Search tool (Media Suite version 3, July 2018).

'humanities', or even wrong, like 'Early modern history' – this appears to be a bit too far in the past for a video interview collection (see the results list in Figure 3).

The search results (i.e. descriptions of individual interviews) can be easily selected to save and export to the personal user space in the Media Suite. This is a restricted environment that is only accessible with a personal login, where a user can store different query results (corpora). Notably, it is only the selected items' metadata that can be stored in the user space; the actual data – a transcript, a video interview, a documentary, et cetera, remain 'in the collection'. Conveniently, the selected items in the user space can be accessed directly by clicking on 'View in catalogue'. In case of oral history data, this is quite laborious: The user is redirected to the DANS website, where, after a login, and depending on the access status of the particular interview, one can watch video interviews or download transcripts, if available (see Figure 4).

In conclusion, it has proven rather difficult to get a satisfying overview of the relevant interviews to be found. Moreover, repeating every step verifying the potential usefulness of all 655

Figure 4. Screenshot of the DANS interface for accessing “datasets” (Sept. 2018).

(and perhaps even more) interviews in the search results would be a laborious process. More importantly, getting access to the content itself is a very challenging issue. For the pilot project, I decided to pragmatically limit myself to the ‘getuigenverhalen’ (‘witness stories’) collection, which contains about 500 interviews on different aspects of World War II. This provided me with a relevant, demarcated, diverse and above all (mostly) open access oral history interview collection to work with. This collection is hosted by DANS and is viewable in the Media Suite.

Newspaper articles

Whereas oral history collections by definition contain eyewitness accounts, and while they have useful (although far from perfect) metadata about the interview content available, creating a corpus of World War II eyewitness accounts from newspapers from the post-war period is far more difficult and laborious. To do so, a researcher needs to rely on a set of search terms, which are arbitrary and limited. The initial aim to include ‘all’ newspaper content derived from an interview with one or

Figure 5. One of the articles found by using the query in Delpher (from: Het vrije volk, 4 May, 1955).

more World War II eyewitnesses turned out to be too ambitious. Decided was to include ‘interview’, ‘vraaggesprek’ (‘interview’) and ‘gesprek’ (‘conversation’) in the query, because this way the particular articles explicated the role of witnesses as a source of information, just as oral history interviews do. The final query was therefore (tweede wereldoorlog) AND (getuige OR ooggetuige OR overlevende) AND (interview OR vraaggesprek OR gesprek). This returned 492 articles, limited to national newspapers, between 1 September, 1944 and 31 December 1995.

Because the aim of the research was to get a general idea of the content of eyewitness accounts of World War II in the Netherlands since 1945, no specific newspaper titles were selected, but keyword searches were performed in all available newspapers from the post-war period. For pragmatic reasons, regional and local papers were excluded. It might have been helpful if the search could have been limited to article titles, instead of the full text. This was not an option at the time, since the Koninklijke Bibliotheek (KB, ‘National Library of the Netherlands’) newspaper collection had not yet been integrated into the Media Suite, and the collection needed to be explored using the KB’s online search interface Delpher (www.delpher.nl). At the time of the publication of this article, the KB newspaper collection has been made accessible in the Media Suite in the same manner as the oral history data: the metadata can be searched in the Media Suite (in a more thorough way than in Delpher), the actual content is stored elsewhere (DANS, KB), and easy links to the particular items are

provided. But even when limiting the search to article titles would have been possible, the manual correction of the results, which contained multitudinous irrelevant ones, would have been laborious.

The workflow, then, was to check every result for its relevance; select the ‘tekst’ (OCR) option and copy the OCR’ed text of each article into a Word file (keeping track of the particular dates and newspaper titles). After having created a corpus in this way, I partitioned the articles into five-year periods, to allow for comparisons over time.

Television documentaries

Similar to the newspaper corpus, the initial idea was to include as many broadcasts as possible that contained eyewitness accounts about World War II, both on television and on radio. For pragmatic reasons, this was limited to television programmes only. The television collection of the Netherlands Institute for Sound and Vision (NISV) offers quite extensive metadata, that can be searched in, again, the ‘single search tool’ in the Media Suite. The search results could be filtered by ‘genre’, at least for this collection. This enabled selecting documentaries and news shows (‘actualiteitenprogramma’s’), for instance, and deselect genres that were unlikely to contain interviews. The latter was especially relevant with regard to sports programmes. Because ‘war’ is a frequently used metaphor in sports discourse, the result list contained relatively many sports programmes. There was no need, therefore to expand the query with the terms ‘interview’, ‘vraaggesprek’ or ‘gesprek’ as was done with the newspaper corpus.

The search was performed in the programme descriptions, that could be selected as a separate metadata field. This ‘TV-sub-corpus’ indeed consisted of programme descriptions, to be compared with the interview transcripts and newspaper articles, and not of the actual video files or annotations thereof. Ideally, this would have been the transcripts of the particular broadcasts, but the functionality of generating transcripts by automatic speech recognition was not available in the Media Suite at the time. Nonetheless, the descriptions were often very exhaustive, not only mentioning the range of topics addressed, but also the names and functions of the people that were interviewed (very relevant in this case), as well as describing the camera shots. The description of an episode of the documentary *Wereld in oorlog*, broadcast on 3 May 1975, for instance, contains no less than 266 words, and makes clear that the episode addressed ‘the shock

of the German invasion, the growth of the Dutch National Socialist party NSB, (...) the February strike, the slowly increasing German control of society, the persecution of the Jews (...), the Arbeitseinsatz, resistance and German repression, and the liberation of the south' and contained 'interviews about this with (...) Th. Ph. van Raalte, Jewish hider, R. Boas-Koopman, survivor concentration camp Vught (...) ds. R.J. van der Veen, resistance fighter (...).' ¹⁰ For a topic analysis, such a description is very useful.

Querying the descriptions for "tweede wereldoorlog" AND (getuige OR overlevende OR ooggetuige) gave 115 results that were relevant for the most part. After ordering these chronologically, the descriptions per item were manually copied into a World file, partitioned in periods of five years each.

Closing remarks

The Media Suite facilitates comparative research into collections of different media types that, traditionally, are studied separately. Especially for oral history research, this is innovative and promising in two ways at least. First, it potentially stimulates the re-use of interviews, not only by media scholars who might accidentally discover this material previously unknown to them, but by (oral) historians as well. Second, as stated above, oral history research is mainly concerned with interviews that have been created by the researcher her- or himself. The metadata as available in the Media Suite – although often unclear – provide some of the context information necessary for the interpretation of a recorded interview, and therewith, for potential re-use. Third, and most importantly, a video annotation tool is provided that enables a novel way to analyse interviews. This tool was not used in this project and was therefore not a part of this chapter. While in traditional oral history research, the manually created transcript is the source that is actually interpreted, by highlighting, commenting, underlining and connecting parts, the annotation tool centralises the interview as video (or audio), and allows for moving away from the written text. The video interviews can be tagged, commented, segmented, and annotations can be exported for further quantitative and qualitative analysis. Potentially, the Media Suite thus enables a novel approach to oral history interviews.

Nonetheless, CrossEWT was not a typical oral history project, both in terms of scale, and regarding its focus on textual sources. The Media Suite, contrarily, is mainly oriented on audio-visual sources. The fact that text analysis tasks could not be performed in the Media Suite is in itself not problematic, only that large amounts of text cannot always be processed by accessible text mining tools. Furthermore, the problem of unclear metadata, both in terms content and regarding their presentation, needs to be addressed in order to enable researchers to adequately and satisfactorily work with the collections that the Media Suite gives access to.

Notes

1. Anette Wieviorka, *The Era of the Witness*. Translated by Jared Stark (Ithaca, NY: Cornell University Press, 2006).
2. Judith Keilbach, "The Eichmann Trial on East German Television – On (not) Reporting about a Transnational Media Event," *VIEW. Journal of European Television History and Culture*, 3, nr. 5 (2014): 17-22, 17.
3. James E. Young, *Writing and Rewriting the Holocaust: Narrative and the Consequences of Interpretation* (Bloomington: Indiana University Press 1998), 132.
4. See for instance Tanja Schult and Diana Popescu eds., *Revisiting Holocaust Representation in the Post-Witness Era* (London: Palgrave MacMillan 2015). See also Susan Hogervorst, "The Era of the User. Testimonies in the Digital Age," *Rethinking History* 24, nr. 2 (2020): 169-183, doi: 10.1080/13642529.2020.1757333.
5. See for a historiographical overview: Malin Thor Tureby, "To Hear with the Collection: The Contextualisation and Recontextualisation of Archived Interviews," *Oral History* 41, nr. 2 (Autumn 2013): 63-74.
6. See for this issue Prue Chamberlayne, Joanna Bornat, and Tom Wengraf eds., *The Turn to Biographical Methods in Social Science: Comparative Issues and Examples* (London: Routledge 2000).
7. Stef Scagliola and Franciska de Jong, "Clio's Talkative Daughter Goes Digital. The Interplay Between Technology and Oral Accounts as Historical Data" in: *The Making of the Humanities, Volume III: The Modern Humanities*, eds. Rens Bod, Jaap Maat and Thijs Weststeijn (Amsterdam: Amsterdam University Press, 2014), 511-526.
8. Ron Grele, "Oral History as Evidence" in: *Handbook of Oral History*, eds. Thomas L. Charlton, Lois E. Myers and Rebecca Sharpless (Altamira Press, Lanham et al., 2008), 43-103, 64.
9. Sara Jones, "Mediated Immediacy: Constructing Authentic Testimony in Audio-visual Media", *Rethinking History* 21, nr. 2 (2017): 135-153, 141, doi: 10.1080/13642529.2017.1305726.

10. Aan de orde komen de schok van de Duitse inval, de groei van de NSB en het ontstaan van de Nederlandsche Unie, de ambtelijke medewerking aan de persoonsregistratie, de Februaristaking, de langzaam toenemende Duitse greep op de samenleving, de Jodenvervolgung, de Arbeitseinsatz en onderduikers, het harder wordende verzet en de Duitse repressie, de bevrijding van het zuiden, de spoorwegstaking van 1944 en de Hongerwinter. Interviews hierover met (...)’

Biography

Susan Hogervorst is historian and works as an assistant professor at the Open University of the Netherlands. Her research involves cultural memory, digital humanities, and oral history. In 2018-2019, she conducted an ADAH research project (funded by CLARIAH and NLeSc) on the representation of violence in historical eyewitness accounts. Currently, she writes a handbook on oral history that will be published by Amsterdam University Press.

TMG Journal for Media History

Volume 24 No (1/2)/2021

DOI

<https://dx.doi.org/10.18146/tmg.810>

PUBLISHER

Netherlands Institute for Sound and Vision

COPYRIGHT

Each article is copyrighted © by its author(s) and is published under license from the author(s). When a paper is accepted for publication, authors will be requested to agree with the Creative Commons Attribution 4.0 International License.