Susan Hogervorst, Open Universiteit (presenter); Hennie Brugman, Meertens Instituut; Lars Buitinck, Huygens ING; Marieke van Erp, KNAW Humanities Cluster; Edwin Klijn, NIOD/Netwerk Oorlogsbronnen; Wouter Kouw, Netherlands eScience Center; Martine de Vos, Netherlands eScience Center; Jeroen Willemsen, Open Universiteit

THE EVENT-DETECTION GAP

Manual vs. automatic event detection in historical research

Tools that automatically detect events in text documents may facilitate historical research [1]. Research in Natural Language Processing (NLP) has yielded several tools to automatically extract information from texts, such as Frog [2] and the NewsReader pipeline [3]. These tools detect events by linguistic and semantic text analysis, by identifying nouns and verbs that denote entities and events, and phrases denoting dates. This information can then be represented in an event structure as provided by the Simple Event Model [4]. However, events can be described in many different ways. In the EviDENce project, we evaluate to what extent we can detect events in eyewitness accounts to create a corpus for further historical research.

In our first case study, which will be the focus of our paper, we use transcripts of the "Getuigenverhalen" oral history interviews with Dutch eyewitnesses of the Second World War. We chose these texts for 3 reasons: 1) they have a certain focus in topics; 2) the spoken text passages of the interviewees contain many mentions of persons, locations, and dates; 3) the questions of the interviewers often contain terms on a more abstract level, that correspond to the metadata (summaries, keywords) of these documents (see examples below). The co-occurrence of concepts and specific examples is likely to help refining the automatic event detection.

Our aim is to evaluate current state-of-the-art NLP tools that automatically detect events and corresponding actors, locations and dates. We compare the results of the automatic event detection to a baseline method in which we use a lexicon approach to select relevant events. Both approaches are evaluated against a manually annotated golden standard in a sample of 4 oral history interview transcripts.

Our preliminary analysis brings up 4 core challenges for the NLP tools in detecting event information. Firstly, the texts contain domain specific terms, often from different languages. For example the interview with Theo Eickman contains the term "Arbeitseinsatz" which the tool mistakenly tags as an entity of type location. Secondly, location detection in many NLP pipelines operates at the level of named locations, such as city names, whereas our texts contain location specifications such as "the doctor's house in Oostermeer". Here the system only returns "Oostermeer". Thirdly, the interviewees are also not always as specific in grounding events in time ("after the war") or by naming specific persons ("the boss"), neither of which is detected by the pipeline. Fourthly, the most complex challenge is that events are often not mentioned explicitly, or described in multiple words or sentences (e.g. "heard a loud bang" or "the floor was shaking" instead of "bombing raid"). To detect these, a system needs

to know which verbs, verb phrases or noun phrases denote events (and if possible relevant to the war domain). This is a task that has not yet been resolved.

By evaluating current state-of-the-art NLP tools for automatic event detection applied in historical research, we can explore the specificities of the event detection gap, and provide useful indications on how to improve event-detection software. Additionally, we contribute to developing event detection as a method for historical research. As can be expected, event detection might especially be relevant in historical research on older texts, for which using modern language lexicons can be problematic. In the second phase of our project, we will therefore expand our scope to digitized ego documents from the 16th century till the present, in order to test this hypothesis, and to provide a substantiated contribution to this important methodological discussion in digital history.

References

[1] Ching-man Au Yeung and Adam Jatowt, 'Studying how the past is remembered: towards computational history through large scale text mining', in B. Berendt, A. de Vries, W. Fan, C. Macdonald, I. Ounis, and I. Ruthven (Eds.), *Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM)* (New York 2011) pp. 1231-1240.

[2] Van den Bosch, A., Busser, G.J., Daelemans, W., and Canisius, S., 'An efficient memory-based morphosyntactic tagger and parser for Dutch' in F. van Eynde, P. Dirix, I. Schuurman, and V. Vandeghinste (Eds.), *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting* (Leuven 2017) pp. 99-114.

[3] Kattenberg, M., Beloki, Z., Soroa, A., Artola, X., Fokkens, A. S., Huijgen, P. E. M. & Verstoep, K., 'Two architectures for parallel processing for huge amounts of text', P*roceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC).* Paris: European Language Resources Association (ELRA), 2016.

[4] Van Hage, W. R., Malaisé, V., Segers, R., Hollink, L., & Schreiber, G. (2011) 'Design and use of the Simple Event Model (SEM)',*Web Semantics: Science, Services and Agents on the World Wide Web*, *9*(2), pp. 128-136.