

UNIVERSITY OF TWENTE
DEPARTMENT OF COMPUTER SCIENCE

Topic Modeling in Memories of Elderly

Author:
Judith Esther Zissoldt

1st Supervisor:
dr. K.P. Truong
2nd Supervisor:
dr.ing. G. Englebienne
3rd Supervisor:
D.S. Nazareth, MSc

A thesis submitted for the degree of

MSc of Science

June 30, 2020

Abstract

The purpose of this thesis was testing and evaluating the performance of different topic models at modeling prominent and returning topics in autobiographical memories of older adults. The models were evaluated on their topic coherence and topic variety. The best performing models in terms of coherence were the ones using noun phrases and words of high and low valence as input. The model with the greatest difference between the topics was an N-Gram topic model. The most common topics found were related to subjects like life and family.

Acknowledgements

Thank you to all my friends, family an co-workers for their support and encouragement.

Contents

1	Introduction	5
2	Background and Related Work	7
2.1	Text Types in Topic Modeling	7
2.2	Topic Models	10
3	Data	12
3.1	MEMOA corpus	12
3.2	Rheumatism corpus	12
3.3	Leven als verhaal corpus	13
3.4	LISS-Panel corpus	13
4	Method	14
4.1	Models	14
4.2	Data pre-processing	15
4.3	Evaluation	16
5	Results	17
5.1	Main Results	17
5.2	N-Gram Topic Model	18
5.3	Noun Phrase LDA	18
5.4	Valence LDA	20
5.5	Bi-term Topic Model	20
5.6	Hidden Markov Topic Model	21
6	Conclusion and Discussion	23
A	Appendix	25
	Bibliography	43

List of Figures

List of Tables

2.1	Overview related work	8
4.1	Overview of the Topic Models	14
5.1	Analysis N-Gram LDA using word-ids	18
5.2	Analysis N-Gram LDA using cluster-ids	19
5.3	Analysis Noun Phrase LDA using word-ids	19
5.4	Analysis Noun Phrase LDA using cluster-ids	19
5.5	Analysis Valence LDA using word-ids	20
5.6	Analysis Valence LDA using cluster-ids	20
5.7	Analysis Bi-term Topic Model using word-ids	21
5.8	Analysis Bi-Term Topic Model using cluster-ids	21
5.9	Analysis Hidden Markov Topic Model using word-ids	22
5.10	Analysis Hidden Markov Topic Model using cluster-ids	22
A.1	Results of the N-Gram LDA	25
A.2	Results of the Noun Phrase LDA	26
A.3	Results of the Valence LDA	26
A.4	Results of the Bi-term Topic Model	27
A.5	Results of the Hidden Markov Topic Model	27
A.6	Results of the Uni-Gram LDA using cluster id's	28
A.7	Results of the Hidden Markov Topic Model using cluster id's	28
A.8	Results of the Biterm Topic Model using cluster id's	29
A.9	Results of the Valence LDA using cluster id's	29
A.10	Results of the Noun Phrase LDA using cluster id's	30
A.11	Evaluation of the N-Gram LDA	31
A.12	Evaluation of the Noun Phrase LDA	32
A.13	Evaluation of the Valence LDA	33
A.14	Evaluation of the Bi-term Topic Model	34
A.15	Evaluation of the Hidden Markov Topic Model	35
A.16	Evaluation of the Unigram Topic Model with cluster id's	36
A.17	Evaluation of the Hidden Markov Topic Model with cluster id's	37
A.18	Evaluation of the Valence Topic Model with cluster id's	38
A.19	Evaluation of the Noun Phrase Topic Model with cluster id's	39
A.20	Evaluation of the Biterm Topic Model with cluster id's	40

Chapter 1

Introduction

Around 50 million people worldwide suffer from Dementia [1]. It affects skills which are necessary for people to perform everyday activities such as a decline in memory and problem-solving. One of the interventions often used to treat people suffering from dementia is reminiscence therapy, where people are asked to discuss past events or experiences with one person or a whole group using support material like photos, music etc. This treatment works especially well for people with dementia due to the fact that a person’s autobiographical memories (memories of personal events) do not immediately disappear after onset of dementia but stay intact for a long time [18]. Reminiscence therapy has proven to improve cognitive functions and reduce symptoms of depression in people suffering from dementia and as a result improve the quality of life of people with dementia as well as their caregivers. Additionally, this intervention also leads to an improved understanding of the individual (person with dementia) thereby promoting relationships and helping to achieve person-centred care [44, 18, 7, 40, 41].

Collecting and analyzing those memories which people talk about during reminiscence therapy can help us understand about important events or experiences in peoples lives. However, analyzing those memories can become a challenge the larger the collection gets. To efficiently analyse and learn about memories on a greater scale we need to be able to identify prominent, returning topics in memories. A method which could present itself useful in the automatic analysis of memories to further the understanding of autobiographical memories might be **topic modeling**. This is a class of text analysis methods which can be used to efficiently and objectively identify prominent, returning topics in life narratives [31].

Topic models fall under the category of unsupervised ML algorithms which means that they require no labeling for learning. Still those algorithms are able to identify meaningful categories also called “topic” by using co-occurrence of words within a document. Using this method enables us to summarize peoples’ experience with the purpose of enhancing their episodic memory [4]. Most often those memories are provided during interviews or conversation e.g., in reminiscence therapy and therefore differ from material normally used for topic modeling in terms of content and structure. However, it needs to be mentioned that not much previous work has been done regarding content of memories and topic modeling. In most cases the goal of research is to create a topic model where the the data is only the tool to create such a well working topic model. This is best done with a great amount of data. Therefore, restricting this amount to fulfill certain conditions is counter-productive. Most training corpora used in topic model research contain articles, like newspapers, scientific papers, or extracts from books. Those types of text are more structured and have a certain length compared for example to interviews. In case memories are collected during conversations or interviews, the information might be less structured. They might differ from a carefully thought about written text. The latter is probably a lot more structured which would make it easier to distinguish different topic/life events within a memory. A person telling a story, on the other hand, might jump back and forth because they forgot something in the beginning and want to add it later on. They might also stop mid-sentence or repeat themselves a lot. Everything we just mentioned are challenges one might encounter when doing topic modeling on autobiographical memories.

The purpose of this thesis is to test and evaluate the performance of different topic models at modeling prominent and returning topics in autobiographical memories. As a way to achieve this goal, exploratory research will be conducted concerning topic modeling on more unstructured

texts like spontaneously told stories. By comparing the performance of different topic models on a corpus of autobiographic memories of older adults we will try to answer the following research questions:

1. *What is a suitable topic modelling method for analyzing autobiographical memories concerning their ability of returning useful results?*
2. *What topics can be identified in a corpus consisting of autobiographical memories using unsupervised machine learning techniques?*

The memories used for this research have not been analyzed concerning the topics they might contain and as a result do not have labels which could be used to compare the results of the topic models against. Thus, the performance of the models will be evaluated based on usefulness of the result, not their correctness, because there is no existing ground truth concerning different topics. Usefulness in this context will be defined as providing coherent and new information.

The next chapter will contain the background necessary for this research. It will introduce different approaches of topic modeling and describe the type of corpus which will be used. Chapter 3 will focus on work that has been done in related fields concerning topic modeling on memories or similar types of texts. Chapter 4 describes the corpus used in the research and how the data was pre-processed. The following chapter introduces the models, which will be used during this research based on the information found in the previous section, as well as section concerning the process of evaluating the models. The last two chapters will describe results of the evaluation, followed by conclusion and discussion to answer the research questions.

Chapter 2

Background and Related Work

2.1 Text Types in Topic Modeling

Different types of texts can differ in their level of structure. There are texts such as articles or books, which are highly structured because their authors followed certain rules when writing them and also had the possibility to think about and rewrite and restructure the text to make it as clear as possible to the reader [12, 17, 8, 14, 15, 23]. Most topic modelling has been done on articles or books [5, 10, 3, 6, 11, 39, 42, 46, 47, 48]. [10] for instance used a topic model to analyse American newspaper articles with respect to the effect of Bush's Terror Alerts on the perceived likelihood of an attack and peoples opinion of his overall performance. Other types of text which has become more popular in the field of topic modeling are social media entries. [38] for example found that topic models can yield insight concerning the level of depression of people by using their entries on twitter by focusing especially on use of language.[2] on the other hand modeled topic changes in online posts over time. However, there are different norms for this type of text, which means that they are often less structured compared to excerpts from books or articles and much shorter. The shorter length can have a negative effect on performance of topic models which work under the assumption that the more often words co-occur in the same text, the higher the probability that they belong to the same topic. However, the lack of sufficient word co-occurrence instances in short texts can make topic models less effective [28, 51]. This has come more apparent since research has started to show more interest in texts found on social media and the information it yields.As a result, new models have been developed which are more effective at modeling short texts.

Finally, the category with the lowest level of structure are transcripts of spoken language, like interview or talks. There has not been done a lot of topic modelling on this type of text. [6] used TED transcripts, among other datasets, to train their model. However, TED talks are probably more structured and less spontaneous compared to interviews, due to the time-limit and expectations the speakers face. Many of the narrative life events that we are using for this research were collected during interviews. They are told without any previous preparation-time by the interviewee, but instead spontaneous told memories which he or she remembered after receiving a prompt word or picture. This means that the people were a lot more likely to jump back and forth within the story because they forgot something in the beginning and want to add it later on[12, 17, 8, 14, 15, 23]. Additionally, they might stop mid-sentence and begin again, because they lose their train of thought[12, 17, 8, 14, 15, 23]. Finally, as mentioned before, the amount of words used to describe a memory can differ greatly between people depending on their personality. Some people might talk a lot and describe everything in detail, while others try to keep it short and summarize what happened in a few sentences.

Only some research has been done in related fields: [4], for instance, describes how topic modeling is used to identify the topics of a person's daily social interactions and using them to summarize their day. [24] explores in what way topic modeling could be used for the detection of topics in children's narratives as a way of predicting coherence and language impairment. [16] uses topic modeling to detect cases of cyber-bullying on social networks because it allows for the extraction of high-level themes in the stories. However, only the paper of [4] uses spontaneous text in form of recorded conversations but those texts do not contain autobiographical memories. Therefore, further research in this field is necessary.

Table 2.1 provides an overview of the topic modeling papers referred to in this chapter. Described are the size and content of the corpora, type of topic model(s) and the method of evaluation. This overview shows that, at least in our selection of papers, there is not much variation in the types of text used as input for topic models. Most of the papers used structured texts like newspaper- or scientific articles, law texts, book extracts. Only a few use less structured texts [6] with the TED talk transcripts or [4] and [24] with transcripts of conversation. Finally, the table shows a great variety in evaluation methods. Even though a lot of the papers used either perplexity or topic coherence, other evaluation methods were for example information retrieval or classification accuracy.

Table 2.1: Overview related work

Article	Corpus	Topic model	Performance
[10]	51766 newspaper articles and transcripts of stories	Latent Dirichlet Allocation [9]	No information about performance
[38]	Around 3 million twitter entries from around 2,000 users. Around 600 of those self-identify as suffering from depression	Supervised Latent Dirichlet Allocation (documents are accompanied by labels or values of interest), Supervised Anchor Latent Dirichlet Allocation (anchor words are associated with each topic)	The performance of the models was measured using mean precision and recall scores
[2]	1500 scientific papers, 3190 newswires, 72592 tweeter entries	Streaming Latent Dirichlet Allocation which models topic and word-topic dependencies between consecutive documents	Perplexity was used to measure performance.
[4]	The researcher collected 21 hours with 375 single conversations.	Latent Dirichlet Allocation [9] in combination with Hidden Markov Model to compute topic evolution over time.	Participants were asked to guess who or what should appear in the image recorded in the situation of conversation based on topic that the words were assigned to.
[24]	The corpus consists of transcripts belonging to 118 adolescents.	Latent Dirichlet Allocation [9]	The performance was measured using topic coherence
[46]	150 abstracts psychological reviews and 150 news-group postings	Bi-gram topic model	Measuring information rates of the corpus in bits per word, where fewer bits per word correspond to a better predictive performance.

[36]	Europarl multilingual parallel corpus with 9672 documents which contain the proceedings of the European parliament , JRC-Acquis multilingual parallel corpus with 23545 documents which contains an approximation of the total body of European Union (EU) law, ACL Anthology Reference Corpus with 10921 documents which is a corpus scholarly publications about Computational Linguistics	A collocation topic model using uni-and bi-grams significant improvement in performance compared to the only bi-gram or basic topic models.	Perplexity was used to measure performance.
[47]	1740 scientific papers	N-gram topic model	Information retrieval
[22]	1740 documents	Hidden Markov Topic Model	Perplexity was used to measure performance.
[6]	19056 news articles, 1096 transcriptions of TED talks, 12812 articles from the Wikipedia and PubMed data sets, 5262 documents from the Jane Austen book collection, 10788 document from the Reuters data set	Copula Latent Dirichlet Allocation that considers sentences a coherent segment and Copula Latent Dirichlet Allocation that considers noun phrases a coherent segment	Perplexity was used to measure performance
[5]	132399 articles from a variety of data sets (WikiTrain1, WikiTrain2, PubMedTrain1, PubMedTrain2, Wiki37, Wiki46, PubMed25, PubMed50)	Sentence Latent Dirichlet Allocation	Perplexity was used to measure performance
[26]	12295 phrases from SearchSnippets, 16,407 samples from July 31st, 2012 to August 14, 2012 from StackOverflow, 19448 entries from the BioASQ's official website, 2472 twitter entries in the period from 2011 to 2012 from the blog which tracks the Text RE-trieval Conference (TREC), 11109 news articles from the GoogleNews site, 4834 captions of the PascalFlickr	Bi-term Topic Model	Classification accuracy was used to evaluate document-topic distribution and topic coherence to measure performance on word-topic distribution.
[50]	230,578 twitter entries published in the REC 2011 microblog between January 23rd and February 8th, 2011	Bi-term Topic Model	Classification accuracy was used to evaluate document-topic distribution and topic coherence to measure performance on word-topic distribution.

[51]	508554 news titles	A short text topic model called Word Network Topic Model	Topic coherence is used as a method to evaluate the performance of the model
[48]	1090 abstracts	Bayesian sentence-based topic model	The summarization performance was measured, by counting the candidate summary and another collection of summaries for reference

The following section will provide some background concerning research in the field of topic modeling and examples of different topic modeling approaches.

2.2 Topic Models

Several topic models use the relationships or closeness between words in a text to define topics. The more often words appear in the same context (a pre-defined range), the more likely they are to belong to the same topic [25]. Due to the unstructuredness of story-telling it can be difficult to define context in terms of document/paragraph etc, because different topics might occur in the same document or even paragraph. Considering this trait, using a topic models with probabilistic approaches to analyze the content of the interviews, might lead to the best results. These models all are based on the same idea, that a document (context) is a mixture of different topics. The Probabilistic Latent Semantic Analysis approach, which was developed by [25], was one of the first to use latent topics in textual corpora modelling.

One of the most well known and used approaches for topic modeling is the so called Latent Dirichlet Allocation (LDA) method [9]. It is an unsupervised probabilistic topic modelling, where each topic has a list of words with probabilities of belonging to this topic and each document has a distribution of topics. Thus, when receiving a document, this algorithm will describe the topic distribution of each document (α), the word distribution of each topic (β), the topic distribution for one document (θ) and the word distribution for one topic (θ). Since its development, this topic modelling approach has been widely used and updated to improve its performance. Additionally, LDA has proven to work well on short texts like tweets [49]. However, other versions of the original LDA model have been developed over the years as well to improve the original LDA model and allow for a better performance for a larger variety of texts.

The original LDA model [9] relied only on the bag-of-words assumption where words are seen as independent of each other and order is not taken into consideration. However, research has shown that word order and phrases (combinations) are important to capture meaning of texts and implementing word order and phrases in a topic model leads to a higher accuracy at the task of defining topics[3, 47]. Several models have been proposed following this idea.

The topic model developed by [46] used pairs of words (bigrams) due to the assumption that a word's appearance is dependent on the previous word and its topic. They found that taking word order into account improved the predictive performance and quality of the topics compared to the basic LDA. [36] also consider the added value of using both uni-and bigrams by focusing on the relationship between both. They assume that similar bigrams, which share the same unigrams, probably belong to the same topic. Their experiment showed that the new approach resulted in significant improvement in performance compared to the only bi-gram or basic topic models. [47] presents a n-grams topic model which discovers topics as well as topical phrases based on context. Their model is based on the bigram topic model [46] and the LDA collocation model [21, 43]. [47] found that the information retrieval performance of the n-gram model was significantly better compared to bigram and basic LDA.

[22] developed a model with the assumption that it is more likely for a sentence to only contain one or two topics at once because it is in itself a coherence segment of the text. Thus, going beyond n-grams towards dependencies between words in sentences. [22] implemented a Hidden Markov Topic Model to model the topics of words in the document as a Markov chain. It allowed them to capture local dependency between words and sentences and use them when calculating topic-probabilities, thereby reducing the number of topics a word could be assigned. Perplexity

was used to measure performance and the topic model using dependencies between words within sentences showed better results than basic LDA.

Another model, implemented by [6] considered Noun Phrases, similar to sentences, to be coherent text segments and uses them for topic modeling. The perplexity measures showed that the model with noun phrases had an overall better performance compared to the original LDA model [9]. Finally, due to the increasing interest in social media entries several new short text topic models have been developed as well. One of them is the Bi-term Topic Model which identifies more global patterns of word co-occurrence to find topics in documents[26, 50]. It works on the assumption that

“two words in a bi term share the same topic drawn from a mixture of topics over the whole corpus” [26].

[26] used classification accuracy to evaluate document-topic distribution and topic coherence to measure performance on word-topic distribution. The results show that the Bi-term topic model performs slightly better than the basic LDA. Another short text topic model was developed by [51]. It is based on the idea that in short texts we can expect the word-word space to still have a certain level of density even when word-by-document space does not. It showed that there are more advantages to using a co-occurrence network to let the model learn about topic components compared to a collection of documents. They used topic coherence to analyse the performance of their model and found that it outperforms basic LDA in classifying short and even normal texts. Overall these short text topic models might prove useful for the shorter memories.

The main focus of this research are the dimension of the corpus which will consist of a collection of important memories or life events of older adults. People are talking about their personal experiences which makes the stories unique. This uniqueness not only relates to the content of the memories, but also the way they are told. Some people might describe the memory with a lot of detail, thereby creating a lot of context and long documents. Others, however, might keep it short and only say a few things. The types of words which are used might also vary greatly because people are not having to take into consideration that other or a large population should be able to understand them as it is with newspaper articles. Therefore, the challenge of this research will be to find a topic model which works well enough for all these types of texts. The five most promising models were chosen, based on different characteristics of the corpus. The N-gram Topic Model[47], Noun Phrase Topic Model[6] and Hidden Markov Topic Model[22] were chosen to test what the best segment size would be. The N-Gram, Noun Phrases and Hidden Markov Topic Model all have a slightly different definition of what a coherent segment is. While the N-gram and Noun Phrase Topic Model both only search for dependency between words in sentences, [22] which his Hidden Markov Topic Model saw whole sentences as coherent segments. The Biterm Topic Model[26, 50] was chosen in the consideration of short length documents of the corpus. Finally, a Valence Topic Model was chosen as well. Valence is one of the dimensions typically used to describe emotions, the other one is arousal [32]. Whereas arousal describes the calming or exciting effect of a stimulus, valence describes how positive or negative it is [32]. Memories are in many cases rich with emotion, implying that they contain a high number of words with low and high valence scores [33]. The Valence Topic Model is based on the assumption that a corpus of words with only high and low valence could improve the usefulness of the resulting topics because words which are more neutral would be filtered out.

Chapter 3

Data

The data was collected from four different corpora using interviews and surveys. All studies were done in the context of autobiographical memory retrieval. In case of the interviews the data was later transcribed using automatic speech recognition and/or manual transcription. The whole data consists of memories from 2345 people above the age of 50 from four different corpora. The memories were collected during interviews and surveys. The whole corpus consisted of 2345 documents and a total of 708559 words. Due to the use of unsupervised learning techniques no labeling of the data was needed.

3.1 MEMOA corpus

The project *Emotion Recognition in Dementia* collected memories of emotional experience [33]. The data came from 23 participants (12 female; 11 male) who were between 65 and 86 years old. The interviews itself were conducted at a place that the participants felt comfortable, in most cases it took place in their homes and lasted between half and one and a half hours. The data which will be used was from two sessions. In the one of the sessions participants were given word cues with different valence and were then asked to recall two emotional memories for each cue. The memories should be of an specific event that had only happened once, in a certain moment and not have lasted longer than a day. In the beginning the participants were given two practice cue words with neutral valence (grass; bread). The cue word were always asked in the same order, first the two neutral words and afterwards one sad cue, followed by a happy one. For each cue the participant should describe a memory and provide pictures to accommodate the memory. Before the other session the interviewers created and prepared a life story book with photos and other documents, provided by the participants during the first session, which aroused emotions. A life story book is meant to keep a person's life story through means of personal items. Those life story books were the subject of the interviews in the second session. The interviews of both sessions were recorded by three microphones, a shotgun and two wireless lavalier microphones. The lavalier microphones were used to record the participant and interviewer and the additional shotgun microphone was placed in front of the participants. During the second session the participants were also filmed using a video camera and heart rate, movement and skin-conductance was measured. A sensors to measure heart rate was put on the torso. Skin-conductance was measured with sensors on the participants fingers and movement was measured using a bracelet around the wrist. The interviews were first transcribed with a automatic speech recognition software and the output was afterwards checked manually.

3.2 Rheumatism corpus

10 interviews were selected from a study of people with rheumatism about stories of their life [37]. The selection based on the age of participants, which had to be above 50 years. The interviews were conducted with two groups of each ten people. Those people were selected based on their MHC-SF score (Mental Health Continuum-Short Form), which measure the well-being. Ten participants with a score higher than 4.57 and lower than 3.43 were selected. The average age of the participants was 54 years, with a range of 26 to 80 years. Due to the fact that the final corpus should contain

only memories from people older than 50, everyone younger was filtered and that is why only 10 interviews remained. The interviews were conducted at the participants' homes and lasted about one hour. They were conducted according to the protocol of the Life Story Interview [19, 29] and focused on the most important events as well as high- and low points of that person's life. The participants were asked to describe the chapters, high - and low points, turning points, the future and values in their life. Finally, a questions was asked concerning the role that rheuma plays in their life story. The interviews were recorded using a single microphone and later transcribed manually.

3.3 Leven als verhaal corpus

From the project *Leven als verhaal* we received 180 documents after selecting the memories of people above 50 years of age. To find participants the University of Twente used an advertisement in the PLUS-Magazine. People were provided with a link to an online form, so they could fill it out anytime and anywhere they felt comfortable without supervision. The goal of the research was the reminiscence and sharing of personal stories concerning your own life in relation to well-being. People were asked to describe three personal memories and rate their well-being. Before the participants were asked to describe the memories they received some instructions concerning the criteria of the memories. First of all, the memories should be the kind that the participants would chose to tell to show other people who they were. Secondly, the memories should be related to important subjects playing a role in your life. Additionally, the memories should be alive, which means that the participants should have thought about them often and that the memories still stir up emotions (positive and negative). Finally, the memories should be at least one year old. After receiving those instructions, the participants were asked to describe all three memories in short, creating a sort of title for each memory. Following those short descriptions, the participants had to describe every memory in detail and explain how they influenced them and why they characterized them as a person. The data was collected in form of a survey, so the memories were already in written form and did not need to be transcribed.

3.4 LISS-Panel corpus

The research concerning autobiographical storytelling and mental health across the lifespans from the University of Twente used data collected by the LISS-Panel (Langlopende Internet Studies voor de Sociale wetenschappen), which is a Dutch panel of around 5000 households from all parts of Dutch society. Those households receive an online questionnaire once or twice every month. The University of Twente was allowed to add their research questions to the monthly panel. Those question were the same as used during the *Leven als verhaal* project. The participants were asked to first describe three important memories in short, then in more detail and explain how they influenced them and why they characterized them as a person. Only memories of people older than 50 were selected and 2.294 documents could be used from the LISS-Panel. The data was collected in form of a survey, so the memories were already in written form and did not need to be transcribed.

Chapter 4

Method

The purpose of this thesis is to test and evaluate the performance of different topic models at modeling prominent and returning topics in autobiographical memories. The performance of different topic models on the data, described in the previous chapter, was compared. The following sections will describe those models as well as the method of evaluation.

4.1 Models

The five most promising topic models were selected based on the conclusion in section 2.2 to evaluate their performance on the corpora of autobiographical memories. The N-Gram LDA, Noun Phrase LDA and Hidden Markov Topic Model were used to possibly reduce the number of topics a word could be assigned to [22] by taking the relationship between words in smaller coherent segments like sentences or phrases into account instead of only the whole of a document. The Bi-term Topic Model was chosen to improve performance on shorter documents of the corpora. Finally, the Valence Topic Model was chosen as a way to analyse a different dimension or meta-data of the corpora. The N-Gram LDA, Noun Phrase LDA and Valence LDA used the same n-gram implementation as base but received a different selection of the corpus. The input of the Noun Phrase model for instance only contained the nouns of the corpus, while another model only contained words of a certain valence. An overview of all models with their respective inputs is presented in table 4.1

Table 4.1: Overview of the Topic Models

Topic Model	Input
N-gram Topic Model	All documents of the corpus as lists containing the word-ids of uni- and/or bi-grams
Noun Phrase Topic Model	All documents of the corpus as lists containing the word-ids of nouns and noun phrases
Valence Topic Model	All documents of the corpus as lists containing the word-ids of only words of a certain valence
Biterm Topic Model	All documents of the corpus as lists containing the word-ids of uni-grams
Hidden Markov Topic Model	All documents of the corpus as lists

N-gram Topic Model An topic model using uni- and/or big-grams was implemented, following the idea presented by [47]. This model received as input all documents of the corpus as lists containing the word-ids of uni- and/or bi-grams.

Noun Phrase Topic Model An n-gram topic model using only the nouns and noun phrases of the text was implemented as well. Based on the article of [6] where noun phrases, similar to sentences, were considered to be coherent text segments. All words of the corpus were tagged using the alpino corpus and non-noun phrases were removed from the corpus. This model received as input all documents of the corpus as lists containing the word-ids of nouns and noun phrases.

Valence Topic Model This model received as input all documents of the corpus as lists containing the word-ids of only words of a certain valence. This approach has (to our knowledge) not been

implemented yet. The model is based on the assumption that a corpus of words with only high and low valence could improve the usefulness of the resulting topics because words which are more neutral would be filtered out. The Moors corpus [32] of Dutch words was for the purpose of identifying words with relevant valence. [32] presented norms including valence for 4,300 Dutch words such as nouns, adjectives, adverbs, and verbs. To identify a word's valence it was rated on a 7-point Likert scale by independent groups of students [32]. Those valence ratings were used for our research to filter the words with a valence higher than 4.5 and lower than 1.5 from the document corpus.

Biterm Topic Model The topic model was implemented based on the articles of [26, 27, 50] chosen in the consideration of short length documents of the corpus. It works on the assumption that *two words in a bi term share the same topic drawn from a mixture of topics over the whole corpus*. This model received as input all documents of the corpus as lists containing the word-ids of uni-grams.

Hidden Markov Topic Model A topic model was implemented which modeled the topics of words in the document as a Markov chain. It allowed them to capture local dependency between words and sentences and use them when calculating topic-probabilities, thereby reducing the number of topics a word could be assigned. This model received as input all documents of the corpus as lists. Those lists contained lists for each sentence in a document with the corresponding word-id of each unigram.

Each model was build twice using different input. The first time, all models received the original text as input. The second time cluster-ids of groups of words with similar meanings (synonyms) were used for the learning process instead of the words (word-ids) themselves. The word embeddings of the Sonar500 [45] were clustered using the K-means method. The clusters were then used to convert the words of our memory corpus to the corresponding cluster-ids. For example, each word of our corpus which appeared in cluster x would receive the cluster-id x. Due to the fact that the embeddings only contained uni-grams and the results from the first iteration showed only a small number of n-grams, the decision was made to only use uni-grams in the second iteration. The optimal number of clusters was determined using the Bayesian Information Criterion error in combination with the elbow method. The figure did not show a distinctive elbow or cutoff point. This result might indicate that there is not an optimal number of clusters, but that a range of clusters-numbers. This range seems to be between 20 and 30 clusters and in the end a decision was made to use 25 clusters. This seemed a very small number, considering what we would like to achieve with the clustering. The clusters should only contain synonyms, different words with similar or the same meaning. Assuming that most words do not have more than 10 synonyms, it can be concluded that k-means should look for more smaller clusters. Therefore another method was used as well to determine the optimal number of clusters. The data we used consisted of 130980 words. Assuming that each word has an average of five synonyms, the number of clusters was defined as 26200.

4.2 Data pre-processing

All memories of one person were defined as a single LDA-document, which were converted to .txt files. Some of the models require seperate files for each document, while others allow for one file in which each document is paragraph. Finally, to prevent problems for the models, spaces were added before and after special signs (.,!?!;). Data cleaning and preprocessing contained the following steps:

1. Stopwords were removed from the text using the Dutch stopwords list from the nltk library [35].
2. All words which contained less than 4 letters were removed.
3. The remaining words were stemmed using the Dutch Snowball stemmer [20].
4. For all the topic models which used n-grams, the Phrase module of the gensim library was used to detect common phrases in the corpus.
5. A dictionary was created from this corpus with a word-id as key and the corresponding word as value.
6. Using the dictionary all words of the corpus were converted to their corresponding word-id/cluster-id.

4.3 Evaluation

To answer the first research question about the pros and cons of different topic models concerning their ability of returning useful results different methods were used. One of those evaluation methods was topic coherence to determine the degree of coherence, meaningfulness and interpretability of the topics that were created. There are several ways in which this can be achieved.

[34] defines two characteristics with which coherence in this context can be measured. The first is the ability of a topic in retrieving documents about a particular subject. The second is the difficulty or ease of coming up with a short label to describe a topic [34, 30]. In most cases only the top 10 terms are provided because they provide enough information to identify the subject of any topic [34].

Each model returned a list with 10 topics, containing 10 words each. The evaluation was done by four raters, who were instructed to evaluate the topics based on their internal coherence [34] and their distinctiveness towards the other topics. Both are necessary to create a sufficient picture of the overall usefulness of the topics. Coherence reflects the internal consistency of a topic and gives an indication of whether the words of a topic relate to each other. To measure coherence, the users had to score each topic on a 3-point scale, designed by [34], where 3 was defined as useful or coherent and 1 as useless/ less coherent. Distinctiveness is a dimension which has not been used yet in the evaluation of topic models. The decision was made to add the indicator after different topic models returned topics with no great variety between them. The topics of the Noun Phrase and Valence Topic Model (Table 5.3, 5.5) are good examples in this case. Even though coherence of each topic was high, the topics themselves looked all very similar, which had a negative effect on the overall usefulness of the topic model due to the lack of new information. The new parameter was meant to show another feature of the returning topics and indicate where such high similarity between topics occurred. Thereby allowing for a broader picture of the resulting topics. Distinctiveness of the topics was also based on a 3-point scale where 3 was defined as different from the rest, 2 as having some similarities with other topics and 1 as being same as another topic. For the results of the run where cluster id's had been used instead of words, the raters were also provided with an excel sheet of those clusters and the corresponding id. This sheet only contained the clusters which could also be found in the topic lists.

To answer the second research question as to what topics could be identified in our corpus of autobiographical memories, the 4 raters were instructed to come up with short labels to describe a topic [34, 30]. However, this was only necessary for those topics which were determined to be useful and were different from the other topics or only had some similarities with them. The raters received all instructions and material (results, file to fill in evaluation) necessary for the evaluation in an e-mail.

An average coherence and difference score as well as standard deviation was calculated for each topic.

Chapter 5

Results

5.1 Main Results

The topics were evaluated based on their internal coherence/usefulness (Coherence score) and their distinctiveness (Difference score) towards the other topics. The scores went from 1 (no coherence/distinctiveness) to 3 (coherent/distinctive). Furthermore, labels were used to describe a topic if those topics were determined to be useful and were different from the other topics or only had some similarities with them. The most common topics found with the topic models using word-id's were related to subjects like life (growing up, childhood, birth, death) and family (relationship, home, parents, children). Topics found by the models using cluster id's were related to economics like finance or business. However, for most of the topics only one or in a few cases two of the raters provided a label.

Three of the models were partly rated by only three instead of four raters. One of the raters did not rate some of the topics on coherence and difference. Those cases were marked by a star (*) in the tables 5.1 to 5.10.

Charts 5.1 and 5.2 summarize the overall results of the different topic models. The blue colored bars represent the topic models using word-id's and the red ones cluster-id's. Chart 5.1 shows the average coherence score of the five topic models. The Valence LDA model (using word-id's) is the model with the highest overall coherence scores of 2.25, but at the same time the highest similarity between topics (Chart 5.2). The Noun Phrase LDA using cluster id's has the overall lowest coherence score of 1 (Chart 5.1). The model with the most topic variety was the Bi-term Topic model using cluster id's with an average difference score of 2.5. The topic model with the lowest distinctiveness between topics was the Noun Phrase Topic Model using cluster-id's. The topic models using cluster-ids did not return a list of words for each topic but instead a list of cluster-ids. Therefore, to evaluate the results, an additional dictionary containing a list of words for each cluster had to be used. The following section will describe the results of each model in more detail.

Chart 5.1: Average coherence score of the topic models

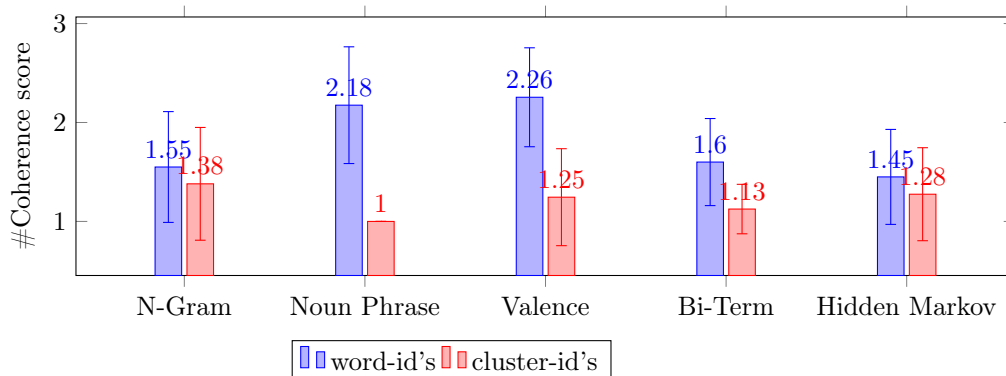
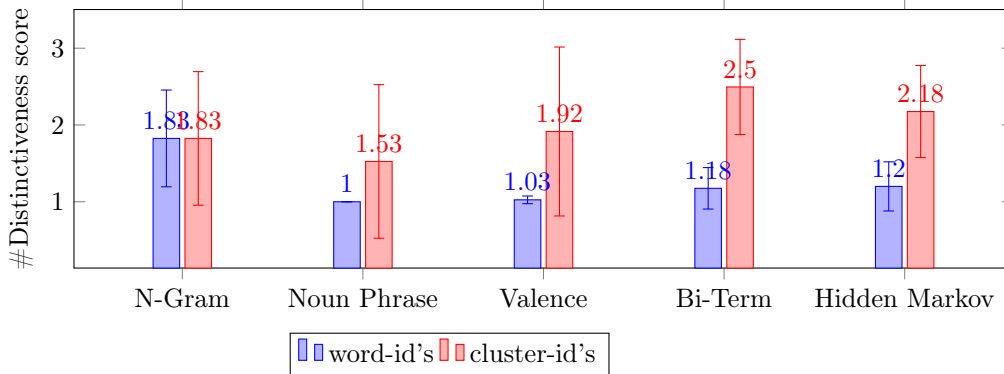


Chart 5.2: Average distinctiveness score of the topic models



5.2 N-Gram Topic Model

Word-ids Topic 7 (Table 5.1) has on average the highest coherence score of 2.5, with 3 being the highest coherence score and 1 being the lowest. The coherence values of topic 3 (5.1) show the highest level of disagreement between the raters with a standard deviation of around 0.957. Interestingly one of the raters gave it a coherence score of 1 but also the short label 'studying' which is very similar to the label 'education' given by another rater who scored the topic as coherent. As mentioned before, this model has the highest difference score of all the model, which can be interpreted as having the highest topic variety. According to the labels of the topics with a coherence-score higher than 1 the raters could identify topics concerning Family, Education, Birth and Death.

Table 5.1: Analysis N-Gram LDA using word-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	1.25	0.5	1.5	0.577
Topic 2	1.25	0.5	1.5	0.577
Topic 3	1.75	0.957	3	0
Topic 4	1.75	0.5	2.25	0.957
Topic 5	1.25	0.5	1.5	1
Topic 6	1.5	0.577	1.5	0.577
Topic 7	2.5	0.577	1.75	0.5
Topic 8	1.25	0.5	2.25	0.957
Topic 9	1.75	0.5	1.5	0.577
Topic 10	1.25	0.5	1.5	0.577

Cluster-ids The average coherence varies between 1 and 1.75 between the different topics (Table 5.2). Four of the topics show a disagreement between the raters concerning the coherence of the topics with a standard deviation of 1 for these topics. 8 of the topics received a label by at least one of the raters, five received one by two. The labels referred to finance, work and business. Topic 6 received the label 'aging' by one of the raters, while the other gave it the label 'vacation & work'.

5.3 Noun Phrase LDA

Word-ids The column containing the short labels (Table A.7 in Appendix A.12) shows that the raters had a high inter-rater agreement concerning the topics, which all contain the word *Family*. The first and third rater got into more detail in the description of the topic and added the descriptors *Life/Birth* and *Death* to their label. All topics received an average coherence score between 2 and 2.25. However, due to the fact that the different topics seemed to be very similar to each other, visible in the similarity of the labels, the topics only received a average different score of 1 (Table 5.3).

Table 5.2: Analysis N-Gram LDA using cluster-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	1.25	0.5	1.75	0.5
Topic 2	1	0	1.75	1
Topic 3	1.5	1	2.25	1
Topic 4	1.75	1	2	0.8
Topic 5	1.75	1	2	0.8
Topic 6	1.75	1	2	1.2
Topic 7	1.3*	0.6*	1.3*	0.6*
Topic 8	1.5	0.6	1.5	0.6
Topic 9	1*	0*	1.7*	1.2*
Topic 10	1*	0*	2*	1*

Table 5.3: Analysis Noun Phrase LDA using word-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	2.25	0.5	1	0
Topic 2	2.25	0.5	1	0
Topic 3	2.25	0.5	1	0
Topic 4	2.25	0.5	1	0
Topic 5	2	0.816	1	0
Topic 6	2.25	0.5	1	0
Topic 7	2.25	0.5	1	0
Topic 8	2	0.816	1	0
Topic 9	2.25	0.5	1	0
Topic 10	2	0.816	1	0

Cluster-ids The fact that the different topics seemed to be very similar to each other is visible in the similarity of the labels and because the topics only received a average different score of around 1.5 (Table 5.4). However, there seems to be a certain level of disagreement between the raters, visible in a standard deviation between 0.96 and 1 for the average difference-score. Three of the four raters assigned an average difference-score between 1 and 1.1, while one rater chose a score of 3 for all the topics. All topics received a coherence score of 1 by all raters, which indicates a high agreement and a no recognizable topics.

Table 5.4: Analysis Noun Phrase LDA using cluster-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	1*	0*	1.75	0.96
Topic 2	1*	0*	1.5	1
Topic 3	1*	0*	1.5	1
Topic 4	1*	0*	1.5	1
Topic 5	1*	0*	1.5	1
Topic 6	1*	0*	1.5	1
Topic 7	1*	0*	1.5	1
Topic 8	1*	0*	1.5	1
Topic 9	1*	0*	1.5	1
Topic 10	1*	0*	1.5	1

5.4 Valence LDA

Word-ids Similar to the Noun Phrase LDA the column containing the short labels (Table A.8 in Appendix A.13) shows that the raters had a higher level of agreement concerning the topics. All labels contain the word *Family* and related words like *Home*. All topics received an average coherence score between 2.25 and 2.3. However, due to the fact that the different topics seemed to be very similar to each other, visible in the similarity of the labels, the topics only received an average difference score between 1 and 1.25. An issue which occurred was that the list of valence-words which was used did not contain all the words of the corpus. Some of those words were also 'pass away' etc which should have a low valence and are therefore interesting due to their low valence and because they appeared quite often in the emotional memories (Table 5.5).

Table 5.5: Analysis Valence LDA using word-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	2.25	0.5	1	0
Topic 2	2.25	0.5	1	0
Topic 3	2.25	0.5	1	0
Topic 4	2.25	0.5	1	0
Topic 5	2.25	0.5	1	0
Topic 6	2.25	0.5	1	0
Topic 7	2.25	0.5	1	0
Topic 8	2.25	0.5	1	0
Topic 9	2.25	0.5	1	0
Topic 10	2.3	0.5	1.25	0.5

Cluster-ids The coherence ratings given to the topics of the Valence LDA using cluster id's range between 1 and 1.5 which indicates a low coherence (Table 5.6). Topic distinctiveness was scored only slightly higher with an average of 1.9. The topic with the highest coherence (1.5) and difference (2.5) score is Topic 10. The labels which were given to the topics for example 'international', 'work' and 'aging'. However, only two of the the raters did give any labels and all of the topics received not more than one.

Table 5.6: Analysis Valence LDA using cluster-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	1.3*	0.6*	2*	1*
Topic 2	1.3*	0.6*	2*	1*
Topic 3	1*	0*	2*	1*
Topic 4	1.3*	0.6*	1.7*	1.5*
Topic 5	1.3*	0.6*	1.7*	1.5*
Topic 6	1*	0*	2*	1*
Topic 7	1.25	0.5	1.75	1
Topic 8	1.25	0.5	1.75	1
Topic 9	1.25	0.5	1.75	1
Topic 10	1.5	1	2.5	1

5.5 Bi-term Topic Model

Word-ids The topics 1 to 3 have the highest average coherence score of 2.25 (Table 5.7). The highest disagreement is about topic 10, where one rater scored 3 while the others scored 1. The topics seem to be very similar to each other because the average difference score for each topic only ranges from 1 to 1.5. The raters identified topics concerning Family and Death. Not all texts are short, there is a range of different length and the question is if the short text might still be too long for this model because it is made for social media entries.

Table 5.7: Analysis Bi-term Topic Model using word-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	2.25	0.6	1.25	0.5
Topic 2	2.25	0.5	1	0
Topic 3	2.25	0.5	1.25	0.5
Topic 4	1.75	0.957	1.25	0.5
Topic 5	2	0.816	1	0
Topic 6	1	0	1.5	0.577
Topic 7	1	0	1.5	0.577
Topic 8	1	0	1	0
Topic 9	1	0	1	0
Topic 10	1.5	1	1	0

Cluster-ids The overall coherence score of the topics of this model is 1.125 and thus low (Table 5.8). It indicates that the raters could not identify a common theme in any topics. There is also little little disagreement between the raters with respect to this. The standard deviation of the coherence score is zero for most of the topics. The topics were rated to be noticeably different from each other with a different scores between 2 and 2.75 . Only one of the raters provided labels for the topics even though only one of the topics was rated higher than 1. Thus the rater provided labels for topics which were rated as being not coherent by them. Only one rater also provided short labels for five of the topics ('finance', 'German', 'politics', 'business', 'family stuff').

Table 5.8: Analysis Bi-Term Topic Model using cluster-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	1.5	1	2.75	0.5
Topic 2	1	0	2.75	0.5
Topic 3	1	0	2.7	0.6
Topic 4	1	0	2.25	1
Topic 5	1	0	2	0.8
Topic 6	1	0	2.5	0.6
Topic 7	1.25	0.5	2.75	0.5
Topic 8	1	0	2.5	0.6
Topic 9	1	0	2.25	0.5
Topic 10	1.5	1	2.5	0.6

5.6 Hidden Markov Topic Model

Word-id's Topic 4 and 6 have share the highest coherence score of 2 and the topics seem to be very similar to each other because the average difference score for each topic only ranges from 0.5 to 1.25 . The labels given to topics 4 and 6 are all very similar and related to the subject of family (Table 5.9).

Cluster-id's The topics of this model given low coherence score between 1 and 1.5 (Table 5.10). The topics were rated to be noticeably different from each other with an average different score of all topics just above 2. Both the coherence score and the difference score have an average standard deviation of around 0.5 , which indicate a high agreement between the raters concerning topic coherence and topic variety. Only one of the topics received a label by two raters and those label do not share a strong connection. One of them is 'government' and the other one 'job/work'.

Table 5.9: Analysis Hidden Markov Topic Model using word-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	1	0	1	0
Topic 2	1.25	0.5	1	0
Topic 3	1.25	0.5	1	0
Topic 4	2	0.816	1.25	0.5
Topic 5	1.5	0.577	1.25	0.5
Topic 6	2	0.816	1.25	0.5
Topic 7	1.5	0.577	1.25	0.5
Topic 8	1	0	1	0
Topic 9	1.75	0.5	1.5	0.577
Topic 10	1.25	0.5	1.5	0.577

Table 5.10: Analysis Hidden Markov Topic Model using cluster-ids

Topic ID	Average Coherence score	SD Coherence	Average Difference score	SD Difference
Topic 1	1.5	0.6	2	0.8
Topic 2	1.5	1	2	0.8
Topic 3	1.5	0.6	2.25	0.5
Topic 4	1	0	2	0.8
Topic 5	1.25	0.5	2.25	0.5
Topic 6	1.5	1	2.25	0.5
Topic 7	1.25	0.5	2.25	0.5
Topic 8	1.25	0.5	2.5	0.6
Topic 9	1	0	2.25	0.5
Topic 10	1	0	2	0.8

Chapter 6

Conclusion and Discussion

Five types of topic models evaluated on their performance at modeling prominent and returning topics in autobiographical memories of older adults. Those were the N-Gram, Noun Phrase, Valence, Bi-Term and Hidden Markov Topic Model.

To answer the first research question *concerning the best suitable topic modelling method for analyzing autobiographical memories in their ability of returning useful results when analyzing a corpus of autobiographical memories* several aspects need to be taken into consideration. The results of the analysis in the previous chapter indicated that the models using word-ids of words with low and high valence (Valence Topic Model) provided topics with the highest coherence. The topics of the Noun Phrase Topic Model had a similar high coherence. Both these models produced topics where the words showed a clear connection, which made it also easier to come up with a title for each topic. However, besides using topic coherence to answer the first question, we also measured difference between the topics. Both the Noun Phrase and Valence topic model returned topics with high coherence but no variety. Essentially only returning a single topic. The topic model with the highest variety of topics was the N-Gram LDA. The Hidden-Markov-Topic-Model and Biterm Topic Model had both low performance on coherence and difference. However, it is probably too early to proclaim that they are not suitable for such a corpus. Their performance might improve on a larger corpus. Overall, it would be interesting to see if the Valence and Noun Phrase approach would keep their status at the two best performing models if there would be more data. N-Gram LDA might improve more than those two with more data because it would be able to use all the words. Valence LDA on the other hand might be able to analyze a new emotional dimension in the data, which the others would not be able to do and those results could be valuable in the scope of therapies.

The second research question concerned *the types of topics which could be identified in a corpus consisting of autobiographical memories using unsupervised machine learning techniques*. The most common topics found were related to subjects like life and family. Those labels are very broad and not as detailed as expected and it reflects the overall level of coherence of the topics (see section 5.1 chart 5.1). We expected topics to be a lot more specific in their focus to represent the different types of memories which people described. Low coherence and broad topics might have presented a challenge for the raters when it came to labelling the data.

The main limitation of this work is the size of the corpora available. The amount of data was not large enough which resulted in undistinctive or incoherent topics. The availability of a larger corpora size would have allowed for stronger conclusions. Therefore an important step for future research would be to add to the corpus of autobiographical memories and collect more through the means of interviews instead of surveys. This allows for questions from the interviewer and might lead to more in-depth's and longer narratives. As a consequence of having less structured types of texts such as interviews, more topic models should be trained on transcripts of spoken language. That way they are also equipped to deal with less structured types of texts such as interviews.

This research used data which was collected through interviews as well as surveys. The choice to use data from two different sources (written text and transcribed spoken language) was made to create a larger corpus and improve the performance of topic models. However, this might have impacted the final results because the topic models had to learn features from two different types of texts, instead of focusing on only one.

Another challenge was related to the Valence Model, which had to use a somewhat limited list

of words with valence ratings provided by [32]. The list of valence-words which was used did not contain all the words of the corpus. Instead of 'pass away' the list for example only contained the word 'dying', which is a synonym that occurred a lot less often in the corpus. Therefore, the model missed out on words with relevant valence which affected the results. To solve this issue the idea of using clusters of words with similar meaning instead of single words as a corpus came up. However, the performance of the topic models using those clusters was overall low in terms of coherence and difference. This was due to the fact that the unsupervised kmeans clustering did not work as intended. The original idea was to set the number of clusters high enough, so that the average number of words per cluster would be around 5. However, that did not work out as expected and the final result was a high number of clusters with only one word and a small number with a lot. Those bigger, less coherent clusters were another limitation of this work. It was difficult for the raters to summarize the big clusters into a word to represent this cluster in the topics. The high number of words in the clusters provided the raters with a great variety of words to lay their focus on. Most of the topics were only given a label (short title) by one, sometimes two raters. This indicates that for many topics half or two thirds of the raters did not think them coherent enough to identify a topic and it is questionable how reliable those labels are. This was probably due to the fact that most of topics consisted only of those clusters which contained a large amount of words themselves. In those case the degree of coherence within each of the clusters was not high, which made it in return difficult to find a coherence between the clusters and within the topic. In some cases when a topic was given a label by two raters they did not refer to a similar or the same theme. An example is the topic 6 of the Unigram Topic Model with cluster id's where one rater gave the label *aging* and another *vacation & work*. The chance of a word being in one of the bigger clusters was just higher. As a result the topics returned by the topic models did not contain cluster id's representing short list of synonyms but instead topics. Still, even though this approach here did not work out well, the idea of using synonyms or small clusters of words instead of single words for topic modeling is still quite interesting and should be investigated further. The method of evaluation, scoring coherence and distinctiveness of topics, could probably remain the same, but the to be evaluated output needs to be improved on. The first step would have to be to find a more effective way of clustering synonyms. Another approach could be using existing synonym databases. However, an issue which might arise with such an approach is distinction of synonym clusters. A word can be a synonym to more than one word, thereby creating a infinitive chain of synonyms.

All in all, it can be said that the current models are not yet really useful in the analysis of memories to help us understand about important events or experiences in peoples lives. The answer to the first research question *which topic model would be best suited for the analysis of autobiographical memories* is inconclusive and the topics which we received were broad or incoherent. Thus, none of the current models would be useful in the task of analyzing the memories which people talk about during reminiscence therapy to help us understand about important events or experiences in peoples lives. However, the findings are still useful for further research. Existing studies (see section 2.1) assessing the quality of topic models in actually identifying topics in data (such as [4, 24, 4, 10]). We also found existing studies that proposed a revised topic model and used the data as a tool to train the model and evaluate performance (such as [2, 38, 46]). This work focused more on what different topic models have to offer for a corpus of memory and what dimensions they are able to show in memories. For further research we recommend to focus on valence as meta-information as is done in the Valence Topic Model. Autobiographical memories contain a high number of words with low and high valence scores [33]. Therefore, making use of this feature of the corpus would mean opening up a new dimension in the data and allowing another understanding of it. This understanding could be useful in the communication with older adults concerning their past and what they experience as emotional events. An improved understanding of the individual would then promote relationships and help to achieve person-centred care.

Appendix A

Appendix

Table A.1: Results of the N-Gram LDA

Topic ID	Words and Probabilities
Topic 1	0.01*eigen, 0.01*gewoon, 0.009*moest, 0.007*allemaal, 0.007*eerst, 0.007*moment, 0.007*helemaal, 0.007*kinder, 0.007*natuurlijk, 0.007*mens
Topic 2	0.007*leven, 0.007*hoofdstuk, 0.006*eigen, 0.006*eerst, 0.006*moment, 0.006*mens, 0.005*ander, 0.005*moeder, 0.005*ding, 0.005*kinder
Topic 3	0.006*niet.weet, 0.004*waanzinn, 0.003*herinnering.weet, 0.003*computer, 0.003*degree, 0.002*mens, 0.002*leven, 0.002*nieuw, 0.002*moest, 0.002*science
Topic 4	0.005*kinder, 0.005*eerst, 0.003*groot, 0.003*puzzel, 0.003*overlijden, 0.003*ouder, 0.003*idee.geen, 0.003*had, 0.003*vader, 0.003*samen
Topic 5	0.006*eerst, 0.005*kinder, 0.004*moest, 0.004*kreeg, 0.004*leven, 0.004*ander, 0.004*femk, 0.003*moment, 0.003*vrouw, 0.003*steeds
Topic 6	0.007*kinder, 0.006*eigen, 0.006*mens, 0.006*ding, 0.005*moment, 0.005*allemaal, 0.004*vrouw, 0.004*nooit, 0.004*moeder, 0.004*gelukkig
Topic 7	0.004*overlijden, 0.003*leven, 0.002*eerst, 0.002*vrouw, 0.002*kinder, 0.002*geborste, 0.002*overleden, 0.002*elkaar, 0.001*gezin, 0.001*ander
Topic 8	0.003*vader, 0.003*eerst, 0.002*moeder, 0.002*noorwegen, 0.002*Arnhem, 0.002*persoon, 0.002*vriend, 0.002*moest, 0.002*scriptie, 0.002*nieuw
Topic 9	0.009*moeder, 0.008*leven, 0.007*kinder, 0.007*moest, 0.007*eigen, 0.007*eerst, 0.006*vader, 0.006*mens, 0.006*natuurlijk, 0.006*verdriet
Topic 10	0.004*kinder, 0.004*hartstikke, 0.004*steeds, 0.004*alleen, 0.004*ding, 0.004*gelukkig, 0.004*mens, 0.004*kreeg, 0.004*eerst, 0.004*moest

Table A.2: Results of the Noun Phrase LDA

Topic ID	Words and Probabilities
Topic 1	0.03*moeder, 0.03*vader, 0.03*lev, 0.03*kinder, 0.02*dochter, 0.02*moment, 0.01*vrouw, 0.01*overlijden, 0.01*geboorte, 0.01*ouder
Topic 2	0.03*moeder, 0.03*kinder, 0.03*vader, 0.03*leven, 0.02*overlijden, 0.02*mens, 0.02*ouder, 0.02*dochter, 0.02*vrouw, 0.01*jaar
Topic 3	0.04*moeder, 0.03*leven, 0.03*vader, 0.03*kinder, 0.02*dochter, 0.02*mens, 0.02*geboorte, 0.02*ouder, 0.02*moment, 0.01*school
Topic 4	0.03*moeder, 0.03*leven, 0.03*vader, 0.03*kinder, 0.02*geboorte, 0.02*ouder, 0.02*mens, 0.02*overlijden, 0.02*moment, 0.01*vriend
Topic 5	0.03*moeder, 0.02*leven, 0.02*vader, 0.02*kinder, 0.02*dochter, 0.02*mens, 0.02*overlijden, 0.01*geboorte, 0.01*jaar, 0.01*ouder
Topic 6	0.03*leven, 0.03*kinder, 0.03*moeder, 0.02*vader, 0.02*mens, 0.02*dochter, 0.02*moment, 0.02*vrouw, 0.02*school, 0.02*overlijden
Topic 7	0.03*leven, 0.03*moeder, 0.02*kinder, 0.02*vader, 0.02*geboorte, 0.02*overlijden, 0.01*moment, 0.01*ouder, 0.01*school, 0.01*vrouw
Topic 8	0.03*vader, 0.03*moeder, 0.02*kinder, 0.02*leven, 0.02*geboorte, 0.02*moment, 0.02*ouder, 0.01*dochter, 0.01*vrouw, 0.01*overlijden
Topic 9	0.03*kinder, 0.03*moeder, 0.03*vader, 0.03*leven, 0.02*vrouw, 0.02*mens, 0.01*dochter, 0.01*geboort, 0.01*moment, 0.01*ouder
Topic 10	0.03*vader, 0.03*kinder, 0.03*leven, 0.02*moeder, 0.02*dochter, 0.02*geboorte, 0.02*mens, 0.02*ouder, 0.01*vrouw, 0.01*moment

Table A.3: Results of the Valence LDA

Topic ID	Words and Probabilities
Topic 1	0.03*leven, 0.03*moeder, 0.03*vader, 0.02*groot, 0.02*school, 0.02*vrouw, 0.01*broer, 0.01*vriend, 0.01*gelukkig, 0.01*dochter
Topic 2	0.05*moeder, 0.04*leven, 0.03*vader, 0.02*groot, 0.02*vrouw, 0.02*natuurlijk, 0.02*school, 0.02*thuis, 0.02*vriend, 0.02*dochter
Topic 3	0.03*vader, 0.03*moeder, 0.02*leven, 0.02*gelukkig, 0.01*groot, 0.01*broer, 0.01*dochter, 0.01*vrouw, 0.01*thuis, 0.01*school
Topic 4	0.03*moeder, 0.02*vader, 0.02*leven, 0.02*groot, 0.01*gelukkig, 0.01*dochter, 0.01*familie, 0.01*vrouw, 0.01*broer, 0.01*thuis
Topic 5	0.03*vader, 0.03*leven, 0.03*moeder, 0.02*groot, 0.02*broer, 0.02*dochter, 0.01*natuurlijk, 0.01*familie, 0.01*school, 0.01*gevoel
Topic 6	0.04*moeder, 0.03*vader, 0.03*leven, 0.02*groot, 0.02*dochter, 0.02*broer, 0.02*natuurlijk, 0.01*vrouw, 0.01*gelukkig, 0.01*eigen
Topic 7	0.02*moeder, 0.02*vader, 0.02*leven, 0.01*broer, 0.01*groot, 0.01*dochter, 0.01*natuurlijk, 0.009*familie, 0.009*thuis, 0.009*school
Topic 8	0.04*moeder, 0.03*vader, 0.03*leven, 0.02*dochter, 0.02*school, 0.02*broer, 0.02*groot, 0.02*vriend, 0.01*vrouw, 0.01*natuurlijk
Topic 9	0.04*moeder, 0.04*vader, 0.03*leven, 0.03*groot, 0.02*vrouw, 0.02*dochter, 0.02*nieuw, 0.02*gelukkig, 0.02*eigen, 0.02*natuurlijk
Topic 10	0.03*moeder, 0.03*vader, 0.03*leven, 0.02*dochter, 0.02*vrouw, 0.02*school, 0.02*groot, 0.02*eigen, 0.01*vakantie, 0.01*gelukkig

Table A.4: Results of the Bi-term Topic Model

Topic ID	Words and Probabilities
Topic 1	0.01*eigen, 0.01*moest, 0.009*moeder, 0.008*vader, 0.007*kinder, 0.007*gewoon, 0.005*allemaal, 0.005*ouder, 0.005*moment, 0.005*nooit
Topic 2	0.01*moeder, 0.009*vader, 0.007*kinder, 0.007*eerst, 0.006*moest, 0.006*ouder, 0.005*herinner, 0.004*had, 0.004*dochter, 0.004*samen
Topic 3	0.01*moeder, 0.01*vader, 0.009*eerst, 0.008*leven, 0.008*kinder, 0.007*dochter, 0.007*overleden, 0.006*overlijden, 0.006*ouder, 0.006*herinner
Topic 4	0.006*leven, 0.006*eerst, 0.005*steeds, 0.004*moest, 0.004*herinner, 0.004*ouder, 0.004*kreeg, 0.004*mens, 0.004*kinder, 0.004*ander
Topic 5	0.007*moest, 0.007*vader, 0.006*kinder, 0.006*moeder, 0.006*eerst, 0.005*ouder, 0.005*herinner, 0.005*mens, 0.004*terug, 0.004*had
Topic 6	0.01*leven, 0.006*kinder, 0.006*mens, 0.006*eerst, 0.006*ding, 0.006*ander, 0.005*steeds, 0.005*nooit, 0.004*groot, 0.004*gewoon
Topic 7	0.008*eerst, 0.007*moest, 0.006*leven, 0.005*kreeg, 0.004*mens, 0.004*moment, 0.004*nooit, 0.004*vrouw, 0.004*steeds, 0.004*moeder
Topic 8	0.006*leven, 0.006*kinder, 0.005*mens, 0.005*herinner, 0.005*eerst, 0.005*steeds, 0.004*nieuw, 0.004*vrouw, 0.004*ander, 0.004*groot
Topic 9	0.008*eigen, 0.008*moest, 0.007*moment, 0.006*allemaal, 0.006*moeder, 0.006*gewoon, 0.005*kreeg, 0.005*helemaal, 0.005*later, 0.005*eerst
Topic 10	0.006*vader, 0.005*eerst, 0.004*later, 0.004*moeder, 0.004*moest, 0.004*nooit, 0.004*kreeg, 0.004*kinder, 0.004*maand, 0.003*ouder

Table A.5: Results of the Hidden Markov Topic Model

Topic ID	Words and Probabilities
Topic 1	0.01*heel, 0.01*wij, 0.01*we, 0.01*ging, 0.01*weer, 0.008*werk, 0.008*kinder, 0.008*moest, 0.008*goed, 0.008*jaar
Topic 2	0.01*heel, 0.008*jaar, 0.007*werk, 0.007*erg, 0.006*eerst, 0.006*goed, 0.006*kwam, 0.005*dag, 0.005*moest, 0.005*tijd
Topic 3	0.02*jaar, 0.01*heel, 0.01*wij, 0.01*kinder, 0.009*goed, 0.009*gaan, 0.008*leven, 0.007*eerst, 0.006*ging, 0.006*moeder
Topic 4	0.02*jaar, 0.01*heel, 0.01*we, 0.01*moeder, 0.01*vader, 0.009*goed, 0.009*leven, 0.009*kinder, 0.007*weer, 0.007*ouder
Topic 5	0.02*jaar, 0.01*heel, 0.01*we, 0.01*goed, 0.007*eerst, 0.007*weer, 0.007*dag, 0.006*kinder, 0.006*moeder, 0.006*ging
Topic 6	0.02*jaar, 0.01*kinder, 0.01*moeder, 0.01*vader, 0.008*leven, 0.007*eerst, 0.007*ouder, 0.007*we, 0.007*zon, 0.006*weer
Topic 7	0.03*heel, 0.01*goed, 0.01*we, 0.008*eigen, 0.008*weet, 0.008*denk, 0.007*moment, 0.007*ding, 0.007*leven, 0.007*jaar
Topic 8	0.02*jaar, 0.02*we, 0.01*heel, 0.008*moest, 0.007*moeder, 0.006*waar, 0.006*kwam, 0.006*dag, 0.005*weer, 0.005*herinner
Topic 9	0.01*we, 0.009*weer, 0.008*vader, 0.008*eerst, 0.007*moeder, 0.006*kwam, 0.006*huis, 0.006*nooit, 0.006*ging, 0.005*kreeg
Topic 10	0.02*jaar, 0.01*we, 0.008*goed, 0.008*ging, 0.008*dag, 0.006*kwam, 0.006*heel, 0.006*leven, 0.005*wij, 0.005*man

Table A.6: Results of the Uni-Gram LDA using cluster id's

Topic ID	Words and Probabilities
Topic 1	12502*0.022, 22932*0.015, 2225*0.013, 20634*0.012, 15897*0.011, 14411*0.010, 5157*0.008, 563*0.007, 5585*0.007, 13821*0.006
Topic 2	2662*0.014, 22932*0.012, 22225*0.010, 12502*0.007, 5585*0.006, 15897*0.006, 5639*0.006, 17593*0.004, 563*0.004, 14410*0.003
Topic 3	14216*0.040, 22932*0.037, 15897*0.030, 22225*0.022, 5585*0.014, 15247*0.013, 14787*0.012, 12502*0.010, 17593*0.009, 563*0.008
Topic 4	22225*0.107, 15897*0.081, 22932*0.055, 5585*0.042, 5639*0.032, 12502*0.027, 2662*0.024, 17593*0.023, 15247*0.020, 1599*0.016
Topic 5	22932*0.092, 15897*0.079, 22225*0.047, 563*0.039, 5585*0.037, 5639*0.033, 2662*0.031, 1599*0.021, 12502*0.019, 14411*0.018
Topic 6	15897*0.096, 22225*0.060, 2662*0.044, 5585*0.044, 22932*0.037, 5639*0.031, 5777*0.022, 12502*0.022, 17593*0.020, 563*0.019
Topic 7	22225*0.106, 22932*0.038, 15897*0.029, 5639*0.022, 5585*0.021, 12502*0.019, 15247*0.017, 17593*0.013, 2662*0.012, 5153*0.010
Topic 8	6732*0.023, 22225*0.008, 12502*0.006, 5639*0.006, 929*0.004, 3853*0.003, 1599*0.002, 22932*0.002, 17495*0.002, 11735*0.002
Topic 9	15897*0.160, 22225*0.085, 22932*0.037, 12502*0.032, 5585*0.031, 5639*0.029, 2662*0.022, 17593*0.018, 14411*0.015, 15247*0.014
Topic 10	22932*0.094, 15897*0.069, 5585*0.038, 22225*0.030, 5639*0.027, 15247*0.023, 2662*0.022, 12502*0.021, 563*0.020, 13821*0.013

Table A.7: Results of the Hidden Markov Topic Model using cluster id's

Topic ID	Words and Probabilities
Topic 1	22932*0.160, 5639*0.108, 1599*0.091, 15247*0.091, 15897*0.084, 8759*0.077, 20575*0.074, 18911* 0.061, 15780*0.045, 216*0.029
Topic 2	5639*0.259, 22932*0.171, 15897*0.162, 20575*0.056, 22225*0.054, 17593*0.037, 679*0.026, 3493*0.019, 5585*0.015, 1525*0.014
Topic 3	5585*0.134, 5639*0.111, 22225*0.086, 18911*0.085, 20575*0.062, 3876*0.052, 2663*0.045, 563*0.042, 17593*0.038, 3584*0.029
Topic 4	20575*0.203, 17593*0.172, 216*0.088, 15897*0.084, 22225*0.050, 3493*0.038, 2759*0.037, 15780*0.030, 679*0.029, 3876*0.029
Topic 5	15897*0.207, 20575*0.152, 2663*0.105, 5639*0.076, 22932*0.069, 929*0.036, 216*0.036, 1599*0.028, 16084*0.024, 1784*0.024
Topic 6	15897*0.53, 18911*0.049, 20585*0.048, 17593*0.038, 3876*0.026, 5636*0.021, 8759*0.020, 20575*0.020, 679*0.019, 22225*0.019
Topic 7	22932*0.282, 20575*0.135, 18911*0.049, 17593*0.048, 22225*0.048, 15897*0.045, 6510*0.044, 16084*0.041, 8759*0.034, 2662*0.021
Topic 8	15897*0.206, 22225*0.153, 8759*0.091, 18911*0.074, 1599*0.073, 20575*0.072, 2663*0.034, 15247*0.033, 4736*0.027, 15780*0.027
Topic 9	5159*0.198, 15897*0.130, 19151*0.095, 22932*0.073, 20575*0.063, 679*0.040, 18911*0.031, 2663*0.030, 5639*0.027, 22225*0.024
Topic 10	15897*0.300, 22932*0.208, 5585*0.073, 563*0.053, 22225*0.042, 216*0.040, 679*0.037, 3876*0.026, 17593*0.025, 20575*0.019

Table A.8: Results of the Biterm Topic Model using cluster id's

Topic ID	Words and Probabilities
Topic 1	16336*0.029, 5585*0.024, 3584*0.024, 563*0.022, 3583*0.021, 17019*0.019, 4736*0.018, 22225*0.018, 12502*0.017, 8604*0.016
Topic 2	17956*0.015, 12419*0.012, 17495*0.012, 4153*0.012, 18438*0.012, 17593*0.010, 706*0.009, 24489*0.008, 17223*0.008, 24036*0.007
Topic 3	23337*0.033, 2172*0.023, 13502*0.023, 17831*0.022, 2994*0.020, 17246*0.020, 12428*0.019, 1552*0.019, 666*0.018, 1791*0.017
Topic 4	17956*0.026, 17495*0.025, 13805*0.020, 17593*0.019, 12419*0.017, 21476*0.017, 24489*0.016, 24*0.016, 10807*0.015, 10711*0.015
Topic 5	10807*0.009, 13805*0.009, 17495*0.008, 17956*0.008, 17593*0.007, 24489*0.007, 10711*0.006, 11659*0.006, 22420*0.006, 2800*0.006
Topic 6	2127*0.007, 25962*0.007, 17223*0.007, 24489*0.006, 2994*0.006, 21760*0.006, 25681*0.006, 6319*0.006, 17246*0.006, 17956*0.005
Topic 7	2058*0.025, 19513*0.023, 17301*0.022, 5982*0.018, 24146*0.017, 2869*0.016, 12428*0.015, 7548*0.015, 11268*0.014, 12289*0.014
Topic 8	11268*0.021, 12428*0.020, 15934*0.020, 24493*0.020, 175*0.016, 2103*0.016, 17301*0.016, 2183*0.014, 3481*0.013, 14536*0.013
Topic 9	12428*0.021, 2994*0.017, 5585*0.017, 17956*0.016, 17495*0.016, 17593*0.016, 23337*0.015, 24489*0.014, 16336*0.013, 17223*0.013
Topic 10	2127*0.025, 25962 0.023, 2994*0.022, 21760*0.022, 7316*0.020, 6319*0.020, 8565*0.019, 5709*0.019, 17223*0.018, 24489*0.018

Table A.9: Results of the Valence LDA using cluster id's

Topic ID	Words and Probabilities
Topic 1	22225*0.215, 5777*0.146, 2662*0.064, 20585*0.057, 25003*0.051, 15247*0.049, 15897*0.042, 2663*0.030, 20985*0.030, 14039*0.029
Topic 2	16336*0.123, 5639*0.112, 15897*0.104, 22225*0.080, 14921*0.075, 5777*0.070, 20585*0.057, 5585*0.035, 13821*0.033, 14411*0.032
Topic 3	15897*0.167, 22932*0.125, 5639*0.081, 22225*0.061, 5777*0.051, 20585*0.045, 14411*0.034, 5585*0.028, 2662*0.027, 5153*0.017
Topic 4	22225*0.203, 15897*0.110, 22932*0.073, 5585*0.051, 5639*0.048, 13821*0.041, 5777*0.036, 14411*0.027, 25003*0.022, 5639*0.020
Topic 5	22225*0.144, 15897*0.101, 22932*0.094, 13821*0.080, 5777*0.078, 5639*0.077, 2663*0.039, 15247*0.034, 2662*0.022, 5585*0.022
Topic 6	15897 0.200, 25003 0.057, 15247 0.051, 22225*0.045, 5585*0.043, 20985*0.040, 12502*0.037, 22932*0.034, 12620*0.033, 1496*0.025
Topic 7	22225*0.155, 15897*0.120, 14411*0.090, 5639*0.071, 22932*0.057, 5585*0.047, 20585*0.034, 5777*0.026, 13434*0.017, 2662*0.015
Topic 8	15897*0.279, 22225*0.076, 5777*0.060, 2662*0.048, 5639*0.040, 22932*0.040, 13434*0.034, 5585*0.032, 13821*0.027, 5153*0.026
Topic 9	15897*0.219, 22225*0.185, 5639*0.079, 14411*0.037, 22932*0.035, 5585*0.031, 5777*0.027, 5153*0.019, 12502*0.018, 2662*0.015
Topic 10	22076*0.142, 13579*0.095, 5585*0.015, 7603*0.015, 13434*0.015, 670*0.015, 15897*0.011, 3897*0.008, 20585*0.008, 2662*0.008

Table A.10: Results of the Noun Phrase LDA using cluster id's

Topic ID	Words and Probabilities
Topic 1	15897*0.154, 22225*0.135, 22932*0.074, 5639*0.057, 2662*0.054, 5585*0.042, 12502*0.037, 15247*0.036, 14411*0.030, 13821*0.016
Topic 2	15897*0.148, 22225*0.096, 22932*0.087, 5639*0.060, 2662*0.042, 12502*0.035, 5585*0.034, 14411*0.025, 15247*0.022, 13821*0.021
Topic 3	22225*0.157, 22932*0.110, 15897*0.100, 2662*0.062, 5639*0.051, 5585*0.041, 12502*0.034, 14411*0.027, 15247*0.024, 13821*0.022
Topic 4	22225*0.144, 15897*0.138, 22932*0.086, 5639*0.074, 2662*0.054, 5585*0.038, 12502*0.034, 15247*0.029, 14411*0.021, 13821*0.020
Topic 5	15897*0.188, 22225*0.120, 22932*0.090, 5639*0.066, 2662*0.046, 5585*0.038, 15247*0.031, 14411*0.027, 12502*0.026, 929*0.017
Topic 6	15897*0.167, 22225*0.129, 22932*0.088, 5639*0.055, 5585*0.045, 2662*0.045, 12502*0.043, 15247*0.024, 14411*0.024, 929*0.020
Topic 7	15897*0.001, 22225*0.001, 22932*0.001, 5639*0.001, 2662*0.001, 12502*0.001, 5585*0.001, 15247*0.001, 14411*0.001, 13821*0.001
Topic 8	22225*0.001, 15897*0.001, 2662*0.001, 22932*0.001, 5639*0.001, 5585*0.001, 12502*0.001, 14411*0.001, 13821*0.001, 15247*0.001
Topic 9	22225*0.001, 15897*0.001, 22932*0.001, 5639*0.001, 5585*0.001, 12502*0.001, 2662*0.001, 15247*0.001, 14411*0.001, 13821*0.001
Topic 10	22225*0.133, 15897*0.125, 22932*0.100, 5639*0.060, 2662*0.058, 5585*0.036, 14411*0.033, 12502*0.031, 15247*0.030, 13821*0.023

Table A.11: Evaluation of the N-Gram LDA

Topic ID	Short label	Coherence score	Difference score
Topic 1	-	1	1
	-	1	2
	Childhood	1	1
	Life	2	2
Topic 2	-	1	1
	-	1	2
	Growing Up	1	1
	Reflection on life	2	2
Topic 3	-	1	3
	Education	3	3
	Studying	1	3
	Education	2	3
Topic 4	Family	2	2
	Childhood	2	3
	First Child	1	1
	Dying	2	3
Topic 5	-	1	1
	Relationships and birth	2	3
	First Child	1	1
	-	1	1
Topic 6	-	1	1
	Birth	2	2
	Giving Birth	1	1
	Children	2	2
Topic 7	Life and death	3	2
	Dying	3	2
	Death of a Family Member	2	2
	Death and birth	2	1
Topic 8	-	1	3
	'From home'	2	3
	Parents	1	1
	-	1	2
Topic 9	Family	2	1
	-	1	2
	Parents	2	1
	Parents	2	2
Topic 10	-	1	1
	-	1	2
	Giving Birth	2	1
	-	1	2

Table A.12: Evaluation of the Noun Phrase LDA

Topic ID	Short label	Coherence score	Difference score
Topic 1	Family, Life and Death	3	1
	Family	2	1
	Family Life	2	1
	Life	2	1
Topic 2	Family, Life and Death	3	1
	Family	2	1
	Family Life	2	1
	Life	2	1
Topic 3	Family, Life and Death	3	1
	Family	2	1
	Death of Family Member	2	1
	Parents	2	1
Topic 4	Family, Life and Death	3	1
	Family	2	1
	Birth of Family/Children	2	1
	Children	2	1
Topic 5	Family, Life and Death	3	1
	Family	2	1
	Family Life	2	1
	-	1	1
Topic 6	Family, Life and Death	3	1
	Family	2	1
	Death of a Family Member	2	1
	Children	2	1
Topic 7	Family, Life and Death	3	1
	Family	2	1
	Family Life	2	1
	Children	2	1
Topic 8	Family, Life and Death	3	1
	Family	2	1
	Family Life	2	1
	-	1	1
Topic 9	Family, Life and Death	3	1
	Family	2	1
	Birth of Family/Children	2	1
	Family	2	1
Topic 10	Family, Life and Death	3	1
	Family	2	1
	Birth of Family/Children	2	1
	-	1	1

Table A.13: Evaluation of the Valence LDA

Topic ID	Short label	Coherence score	Difference score
Topic 1	Family and Home	3	1
	Family	2	1
	Positive Family Life	2	1
	Family	2	1
Topic 2	Family and Home	3	1
	Family	2	1
	Family Life	2	1
	Family	2	1
Topic 3	Family and Home	3	1
	Family	2	1
	Positive Family Life	2	1
	Home	2	1
Topic 4	Family and Home	3	1
	Family	2	1
	Positive Family Life	2	1
	Family	2	1
Topic 5	Family and Home	3	1
	Family	2	1
	Family Life	2	1
	Parents	2	1
Topic 6	Family and Home	3	1
	Family	2	1
	Positive Family Life	2	1
	Family	2	1
Topic 7	Family and Home	3	1
	Family	2	1
	Family Life	2	1
	Family	2	1
Topic 8	Family and Home	3	1
	Family	2	1
	Family Life	2	1
	Family	2	1
Topic 9	Family and Home	3	1
	Family	2	1
	Positive Family Life	2	1
	Family	2	1
Topic 10	Family and Home	3	1
	Family	2	1
	Positive Family Life	2	1
	On vacation with the family	2	2

Table A.14: Evaluation of the Bi-term Topic Model

Topic ID	Short label	Coherence score	Difference score
Topic 1	Family	3	1
	Family	2	1
	Youth	2	1
	Parents	2	2
Topic 2	Family	3	1
	Children	2	1
	Youth	2	1
	Family	2	1
Topic 3	Family and Death	3	1
	Children/Death	2	1
	Death of a Family Member	2	1
	Dying	2	2
Topic 4	Family	3	1
	Children	2	1
	-	1	2
	-	1	1
Topic 5	Family	3	1
	-	1	1
	Youth	2	1
	Children	2	1
Topic 6	-	1	1
	-	1	1
	-	1	2
	-	1	2
Topic 7	-	1	1
	-	1	1
	-	1	2
	-	1	2
Topic 8	-	1	1
	-	1	1
	-	1	1
	-	1	1
Topic 9	-	1	1
	-	1	1
	-	1	1
	-	1	1
Topic 10	Family	3	1
	-	1	1
	-	1	1
	-	1	1

Table A.15: Evaluation of the Hidden Markov Topic Model

Topic ID	Short label	Coherence score	Difference score
Topic 1	-	1	1
	-	1	1
	-	1	-
	-	1	1
Topic 2	Time	2	1
	-	1	1
	-	1	-
	-	1	1
Topic 3	-	1	1
	-	1	1
	-	1	-
	Youth	2	1
Topic 4	Life as family	3	1
	-	1	1
	Family and Work	2	1
	Family	2	2
Topic 5	-	1	1
	-	1	1
	Family and Work	2	1
	Youth	2	2
Topic 6	Life as family	3	1
	-	1	1
	Family	2	1
	Parents	2	2
Topic 7	Time	2	1
	-	1	1
	-	1	1
	Life	2	2
Topic 8	-	1	1
	-	1	1
	-	1	1
	-	1	1
Topic 9	Home	2	1
	-	1	1
	Parents	2	2
	Work	2	2
Topic 10	Time	2	1
	-	1	1
	-	1	2
	-	1	2

Table A.16: Evaluation of the Unigram Topic Model with cluster id's

Topic ID	Short label	Coherence score	Difference score
Topic 1	-	1	1
	-	1	2
	financial	1	2
	work & finance	2	2
Topic 2	-	1	1
	-	1	2
	travel	1	1
Topic 3	-	1	1
	-	1	2
	life	3	3
Topic 4	-	1	1
	-	1	2
	business/sports	3	3
Topic 5	-	1	1
	-	1	2
	sports/business	3	3
Topic 6	-	1	1
	-	1	1
	aging	3	3
Topic 7	-	1	1
	-	1	1
	vacation & work	2	3
Topic 8	-	1	1
	-	1	1
	money & job	-	-
Topic 9	-	1	1
	-	1	1
	international finance	2	2
Topic 10	-	1	1
	-	1	1
	finance & company	2	2
Topic 11	-	1	1
	-	1	1
	-	-	-
Topic 12	-	1	1
	-	1	1
	-	1	3
Topic 13	-	1	1
	-	1	2
	-	-	-
Topic 14	-	1	1
	-	1	2
	-	1	3

Table A.17: Evaluation of the Hidden Markov Topic Model with cluster id's

Topic ID	Short label	Coherence score	Difference score
Topic 1	War - Sports -	2 1 2 1	1 2 2 3
Topic 2	- - Business -	1 1 3 1	1 2 2 3
Topic 3	- - Government job/work	1 1 2 2	2 2 2 3
Topic 4	- - - -	1 1 1 1	2 2 1 3
Topic 5	- - aging/society -	1 1 2 1	2 2 2 3
Topic 6	- - aging/care -	1 1 3 1	2 2 2 3
Topic 7	- - sports -	1 1 2 1	2 2 2 3
Topic 8	- - aging -	1 1 2 1	2 3 2 3
Topic 9	- - - -	1 1 1 1	2 2 2 3
Topic 10	- - - -	1 1 1 1	2 1 2 3

Table A.18: Evaluation of the Valence Topic Model with cluster id's

Topic ID	Short label	Coherence score	Difference score
Topic 1	-	1	1
	-	1	2
	-	-	-
	company & vacation	2	3
Topic 2	-	1	1
	-	1	2
	-	-	-
	american/international	2	3
Topic 3	-	1	1
	-	1	2
	-	-	-
	-	1	3
Topic 4	-	1	1
	-	1	1
	-	-	-
	job/work	2	3
Topic 5	-	1	1
	-	1	2
	-	-	-
	job/work	2	3
Topic 6	-	1	1
	-	1	2
	-	-	-
	-	1	3
Topic 7	-	1	1
	-	1	2
	aging	2	1
	-	1	3
Topic 8	-	1	1
	-	1	2
	aging	2	1
	-	1	3
Topic 9	-	1	1
	-	1	2
	aging	2	1
	-	1	3
Topic 10	-	1	1
	-	1	3
	international	3	3
	-	1	3

Table A.19: Evaluation of the Noun Phrase Topic Model with cluster id's

Topic ID	Short label	Coherence score	Difference score
Topic 1	-	1	1
	-	1	2
	-	-	1
	-	1	3
Topic 2	-	1	1
	-	1	1
	-	-	1
	-	1	3
Topic 3	-	1	1
	-	1	1
	-	-	1
	-	1	3
Topic 4	-	1	1
	-	1	1
	-	-	1
	-	1	3
Topic 5	-	1	1
	-	1	1
	-	-	1
	-	1	3
Topic 6	-	1	1
	-	1	1
	-	-	1
	-	1	3
Topic 7	-	1	1
	-	1	1
	-	-	1
	-	1	3
Topic 8	-	1	1
	-	1	1
	-	-	1
	-	1	3
Topic 9	-	1	1
	-	1	1
	-	-	1
	-	1	3
Topic 10	-	1	1
	-	1	1
	-	-	1
	-	1	3

Table A.20: Evaluation of the Biterm Topic Model with cluster id's

Topic ID	Short label	Coherence score	Difference score
Topic 1	-	1	2
	-	1	3
	finance	3	3
	-	1	3
Topic 2	-	1	2
	-	1	3
	German	1	3
	-	1	3
Topic 3	-	1	2
	-	1	3
	anything	1	-
	-	1	3
Topic 4	-	1	2
	-	1	3
	politics	1	1
	-	1	3
Topic 5	-	1	2
	-	1	2
	business	1	1
	-	1	3
Topic 6	-	1	2
	-	1	3
	undefined	1	2
	-	1	3
Topic 7	-	1	2
	-	1	3
	family stuff	2	3
	-	1	3
Topic 8	-	1	2
	-	1	3
	undefined	1	2
	-	1	3
Topic 9	-	1	2
	-	1	2
	undefined	1	2
	-	1	3
Topic 10	-	1	2
	-	1	3
	undefined	1	2
	-	1	3

Bibliography

- [1] Alzheimer's Disease International. World Alzheimer Report. *Alzheimer's Disease International*, London, UK (2018).
- [2] Amoualian, H., Clausel, M., Gaussier, E., & Amini, M. R. (2016, August). Streaming-lda: A copula-based approach to modeling topic dependencies in document streams. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 695-704). ACM.
- [3] Ashuha, A., & Loukachevitch, N. (2016, April). Bigram Anchor Words Topic Model. In *International Conference on Analysis of Images, Social Networks and Texts* (pp. 121-131). Springer, Cham.
- [4] Bahrainian, S. A., & Crestani, F. (2016, October). Cued retrieval of personal memories of social interactions. In *Proceedings of the first Workshop on Lifelogging Tools and Applications* (pp. 3-12). ACM.
- [5] Balikas, G., Amini, M. R., & Clausel, M. (2016, July). On a topic model for sentences. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 921-924). ACM.
- [6] Balikas, G., Amoualian, H., Clausel, M., Gaussier, E., & Amini, M. R. (2016, December). Modeling topic dependencies in semantically coherent text spans with copulas. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 1767-1776).
- [7] Beste, H. M., & Richardson, R. G. (1981). Developing a Life Story Book Program for Foster Children. *Child Welfare*, 60(8).
- [8] Blankenship, J. (1962). A linguistic analysis of oral and written style. *Quarterly Journal of Speech*, 48(4), 419-422.
- [9] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- [10] Bonilla, T., & Grimmer, J. (2013). Elevated threat levels and decreased expectations: How democracy handles terrorist threats. *Poetics*, 41(6), 650-669.
- [11] Boyd-Graber, J. L., & Blei, D. M. (2009). Syntactic topic models. In *Advances in neural information processing systems* (pp. 185-192).
- [12] Chafe, W., & Tannen, D. (1987). The relation between written and spoken language. *Annual review of anthropology*, 16(1), 383-407.
- [13] A Python multi-threads implementation of Hidden Topic Markov Model. Received from: <https://github.com/Charleo85/pyhtmm>
- [14] DeVito, J. A. (1965). Comprehension factors in oral and written discourse of skilled communicators. *Communications Monographs*, 32(2), 124-128.
- [15] DeVito, J. A. (1967). A linguistic analysis of spoken and written language. *Communication Studies*, 18(2), 81-85.

- [16] Dinakar, K., Jones, B., Lieberman, H., Picard, R., Rose, C., Thoman, M., & Reichart, R. (2012, May). You too?! mixed-initiative lda story matching to help teens in distress. In *Sixth International AAAI Conference on Weblogs and Social Media*.
- [17] Drieman, G. H. (1962). Differences between written and spoken language: An exploratory study. *Acta Psychologica*, *20*, 78-100.
- [18] Elfrink, T. R., Zuidema, S. U., Kunz, M., & Westerhof, G. J. (2017). The effectiveness of creating an online life story book on persons with early dementia and their informal caregivers: a protocol of a randomized controlled trial. *BMC geriatrics*, *17*(1), 95.
- [19] Foley Centre (2001). *Coding autobiographical episodes for themes of agency and communion*. Verkregen op 3 maart 2010 via <http://www.sesp.northwestern.edu/foley>.
- [20] <https://radimrehurek.com/gensim/index.html>
- [21] Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in Semantic Representation. *Psychological Review*, *114*(2), 211-244.
- [22] Gruber, A., Weiss, Y., & Rosen-Zvi, M. (2007, March). Hidden topic markov models. In *Artificial intelligence and statistics* (pp. 163-170).
- [23] Halliday, M. A. (1979). Differences between spoken and written language: Some implications for literacy teaching. In *Proc. 4th Australian reading conference, Adelaide, Australia, 1979* (Vol. 2, pp. 37-52).
- [24] Hassanali, K. N., Liu, Y., & Solorio, T. (2013, August). Using Latent Dirichlet Allocation for child narrative analysis. In *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing* (pp. 111-115).
- [25] Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, *42*(1-2), 177-196.
- [26] Jipeng, Q., Zhenyu, Q., Yun, L., Yunhao, Y., & Xindong, W. (2019). Short Text Topic Modeling Techniques, Applications, and Performance: A Survey. *arXiv preprint arXiv:1904.07695*.
- [27] Bi-Term Topic Model (BTM) for very short texts. Received from: <https://github.com/liguoyu1/python/tree/master/TopicModel/BTM>
- [28] Liu, C. Y., Liu, Z., Li, T., & Xia, B. (2018). Topic Modeling for Noisy Short Texts with Multiple Relations. In *SEKE* (pp. 610-609).
- [29] McAdams (2008). *The life story interview*. The Foley Centre for the Study of Lives, Northwestern University. Verkregen op 3 maart 2010 via <http://www.sesp.northwestern.edu/foley>
- [30] Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Hu, X., Tang, J., Gao, H., & Liu, H. (2013, May). Op-timizing Semantic Coherence in Topic Models. In *Proceedings of EMNLP'11* (pp. 262-272).
- [31] Mohr, J. W., & Bogdanov, P. (2013). Introduction—Topic models: What they are and why they matter. *Poetics*, *41* (6), 545-569
- [32] Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A. L., ... & Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior research methods*, *45*(1), 169-177.
- [33] Nazareth, D., Tournier, E., Leimkötter, S., Janse, E., Heylen, D., Westerhof, G. J., & Truong, K.P. (2019). An acoustic and lexical analysis of emotional valence in spontaneous speech: Autobiographical memory recall in older adults.
- [34] Newman, D., Han Lau, J., Grieser, K., & Baldwin, T. (2010). Automatic Evaluation of Topic Coherence. In *Proceedings of Human Language Technologies: The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 100-108).
- [35] <http://www.nltk.org/howto/stem.html>

- [36] Nokel, M., & Loukachevitch, N. (2015). The method of accounting bigram structure in topical models. *Computational Methods and Programming*.
- [37] Oosterzee, A. F. (2011). *Levensverhalen en geestelijke gezondheid bij reumatische patiënten: geslachts-, leeftijd-en welzijnsgerelateerde verschillen in levensverhalen bij reumatische patiënten met betrekking tot agency, communion, contamination, redemption en copingstrategieën* (Master's thesis, University of Twente).
- [38] Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V. A., & Boyd-Graber, J. (2015). Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (pp. 99-107).
- [39] Riedl, M., & Biemann, C. (2012, July). TopicTiling: a text segmentation algorithm based on LDA. In *Proceedings of ACL 2012 Student Research Workshop* (pp. 37-42). Association for Computational Linguistics.
- [40] Ryan T. & Walker R. (1985) *Making Life Story Books*. British Agencies for Adoption and Fostering, London.
- [41] Ryan T. & Walker R. (1993) *Life Story Work: A Practical Guide to Helping Children Understand their Past*. British Association for Adoption and Fostering, London.
- [42] Shafiei, M. M., & Milios, E. E. (2006, December). Latent Dirichlet co-clustering. In *Sixth International Conference on Data Mining (ICDM'06)* (pp. 542-551). IEEE.
- [43] Steyvers, M., & Griffiths, T. (2005). Matlab topic modeling toolbox 1.3.
- [44] Thompson, R. (2011). Using life story work to enhance care. *Nursing older people*, 23(8).
- [45] Tulkens, S., Emmery, C., & Daelemans, W. (2016). Evaluating unsupervised Dutch word embeddings as a linguistic resource. *arXiv preprint arXiv:1607.00225*.
- [46] Wallach, H.M. (2006). Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning*, (pp. 977-984). ACM.
- [47] Wang, X., McCallum, A., & Wei, X. (2007). Topical n-grams: phrase and topic discovery, with an application to information retrieval. In *Seventh IEEE International Conference on Data Mining, ICDM*, (pp. 697-702). IEE.
- [48] Wang, D., Zhu, S., Li, T., & Gong, Y. (2009, August). Multi-document summarization using sentence-based topic models. In *Proceedings of the ACL-IJCNLP 2009 conference short papers* (pp. 297-300). Association for Computational Linguistics.
- [49] Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010). TwitterRank: Finding topic-sensitive influential twitterers. In *Proc. Int. Conf. on Web Search and Data Mining*, 261-270.
- [50] Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013, May). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1445-1456). ACM.
- [51] Zuo, Y., Zhao, J., & Xu, K. (2016). Word network topic model: a simple but general solution for short and imbalanced texts. *Knowledge and Information Systems*, 48(2), 379-398.