

Computational analysis of Life Books – a probing study

Theo Meder

Abstract

Computational analysis of life stories in so-called Life Books is successful on the level of word frequencies, distribution of themes, gender differences, and sentiments with software like AntConc and LIWC2015, but stylometric analysis and clustering are less successful with Stylo because of the interference by interviewers and editors. A program like AntConc can also be used to track thematic gaps in life stories.

Summary

As soon as "data" turn into "big data", analysis by "close reading" can become a problem: it can become an endless process that eventually the brain of the researcher can no longer get a grip on. Methods of computational humanities can partially solve the problem: various tools can be used to make quantitative analyses of large amounts of text, for example in the field of ethnology or folklore. Various tools may be considered for such text analysis. For example, the program AntConc can be used to study word frequencies, as well as the distribution of concepts across the text. LIWC2015 can be used for sentiment analysis of life stories and show differences between genders (or generations) in telling life stories. Stylo may be used for the stylometric analysis of texts. The usefulness of such programs is tested here on a still relatively small corpus of so-called Life Books – in fact still small enough for close reading and qualitative analysis. However, the intention here is to use the corpus as a pilot to see how valuable the tools can be for a much larger amount of texts that will be added in the near future. In this pilot I want to see to what extent the Life Books can be used for structural analysis, gender differences in narrative style and subject choice, sentiment analysis, recurring themes, distribution of motifs, and perhaps most importantly: thematic gaps. That is to say: which (important) issues are not raised by the storytellers?

The experiment shows that it is possible to do research into narrative structures, although this could be much more refined in terms of events. Stylometric analysis with Stylo of male and female repertoires is rather tricky, because interviewers / editors can (very much) interfere as a filter here. Stylo looks for patterns in the use of function words to determine different styles, but Life Books are just not quoting narrators literally all the time, so in quite some cases linguistical features, like the use of function words, may not originate from the storytellers but from the editors.

On the other hand, sentiment analysis in combination with gender, for example, is possible using LIWC2015: this tool can give a fair representation of emotions, relationships and related motifs in life stories. Furthermore, AntConc proves to be a useful tool to investigate the occurrence and distribution of themes and topics. Research into the lack of certain themes and motifs remains an interesting option as well.

Keywords

Life Books, stylometrics, sentiment analysis, concordance, frequency, computational humanities

As soon as "data" turn into "big data", analysis by "close reading" can become a problem: it can become an endless process that eventually the brain of the researcher can no longer get a grip on. Methods of computational humanities can partially solve the problem: various tools can be used to make quantitative analyses of large amounts of text, for example in the field of ethnology. Various tools can be considered for such text analysis. For example, the program AntConc can be used to study word frequencies, as well as the distribution of concepts across the text. LIWC2015 can be used for sentiment analysis of life stories. And Stylo can be used for the stylometric analysis of texts. The usefulness of such programs is tested here on a still relatively small corpus of so-called Life Books – in fact still small enough for close reading and qualitative analysis. However, the intention here is to use the corpus as a pilot to see how valuable the tools can be for a much larger amount of texts that will be added in the near future. In this pilot I want to see to what extent the Life Books can be used for structural analysis, gender differences in narrative style and subject choice, sentiment analysis, recurring themes, distribution of motifs, and perhaps most importantly: thematic gaps. Which (important) issues are not raised by the storytellers?

A computational pilot

In 2013 the Humanitas Foundation, department Almere, started a new project under the name "Levensboek" (Life Book). Volunteers from the Humanitas Foundation went to interview older people to record their life story. The stories were in many cases recorded in mp3 format, and then transcribed and edited by a volunteer. The end result, printed with photos from the storyteller's collection, or with historical illustrations from the Internet, was printed and published as a limited edition booklet. One of the objectives was to keep the elderly mentally active and to put the grey cells to work, as was the case in earlier years in reminiscence projects in care homes. Another goal was that the life stories could be distributed as booklets with testimonies to interested children and grandchildren, family and friends.

One of the initiators of the Almere Humanitas Foundation contacted Theo Meder with the question whether the booklets are not also interesting for the Meertens Institute, for their archive or for research. The Meertens Institute in Amsterdam is a research institute for the language and culture of daily life in the Netherlands, and is part of the Royal Netherlands Academy of Arts and Sciences (KNAW). The Meertens archive contains a lot of material related to the culture of daily life, and has been collecting on diaries, for instance, for a number of years. The Life Books would be a welcome addition to that. A contract was designed in which the storytellers indicate their agreement with the archiving, and from how many years the booklet can be used for research. Part of the storytellers does not feel like it and does not provide a Life Book, another part finds it no problem and in many cases the Life Book may be used immediately for research. The Meertens Institute also always asks for a digital version of the Life Book for computational research – such a digital version is not always included.

In computational research, automatic structure analysis, research into motifs, stylometry and sentiment analysis can be applied (the possibilities can range from folktales, folksongs and life stories to a collection of blogs or tweets on the Internet, for instance). Gender research could also be carried out: do women talk about other subjects than men? The Meertens Institute then drew up a questionnaire of subjects that ethnologists are particularly interested in: folktales, songs, games, celebrations, rituals, and the use of media. Which stories were told to children, and which songs were sung? What was eaten as breakfast, lunch and dinner? How were Queen's Day, Easter, Sinterklaas (Saint Nicolas), Christmas and New Year's Eve celebrated? What did marriages and funerals look like over the years? Which games were played and which sports were practised, and what were the holidays like? What were the favorite radio and television programs? The volunteering interviewers have sometimes taken note of the list, but in practice the list was often neglected. The questionnaire itself, however, was not designed for computational analysis; it was merely meant to be of help during the interviews.

The first Life Book was festively presented in May 2013 at the Meertens Institute. It concerned the book *Met Hart en Ziel* (With Heart and Soul) by Mrs. Elly IJsendijk. In March 2018, the 50th Life Book was handed over to alderman René Peeters (for education, youth and volunteer work, among others) in Almere.¹ At that time, the Meertens Institute had

¹ <http://www.almeredezeweek.nl/nieuws/1453148-wethouder-krijgt-50e-humanitas-levensboek>

not yet received that many Life Books.² After the obvious success in Almere, other departments of Humanitas started making Life Books too, like Apeldoorn and Zaandam.

I was invited to give a presentation on Thursday 4 July 2019 at the office of Humanitas Almere about the research that we are doing with the Life Books. At that moment the Meertens Institute was mainly storing Life Books in the archive for future research, while being involved in large scale research projects on contemporary feasts and rituals, alternative healing, post-colonialism and international belief legends.³ Perhaps a project on ego documents will be started in a later stage, and then diaries and Life Books will be involved. The collection of Life Books may then have grown to several hundred copies, so that we have access to big data. So to the question “What are you doing with those Life Books now?” I sadly had to answer: nothing yet; we currently only archive them. Nevertheless, I had decided to conduct a pilot on the 19 digital Life Books I had at that time, to see if there was interesting research to do using some analytical, digital tools.

I must emphasize that I am a humanities researcher, and that I am interested in digital humanities, but that I – like many of the humanities colleagues my age – am unable to program or devise algorithms myself. I always have to use already designed, preferably menu-driven tools. In the Life Books I searched for structures, sentiments, styles, themes and distribution of motifs, and used tools for this such as LIWC2015 (Linguistic Inquiry and Word Count 2015), Stylo and AntConc. My research is not standard ethnological or anthropological research, where qualitative research into the culture of daily life (for instance by performing fieldwork, observing and interviewing people) is paramount. My pilot involved looking for quantities and patterns in written texts. Besides, the interviews were conducted by volunteers with no ethnological or anthropological experience. They mainly had open conversations with the storytellers. My pilot is focused on the possibilities of future quantitative research based on big data (which 19 life stories are not).

First I will give some basic information about the Life Books and their narrators, and after that I look at the pros en cons of some computational tools.

Narrators, editors and cleaning up the data

The majority of the life story tellers are born in the '20s and '30s; only a few are from after the Second World War. As a child, some have even experienced the crisis years, and most have witnessed the war. Without exception, those years of war have left a deep impression, according to the testimonials.

What is always a daunting task in the computational analysis of digital texts is cleaning up the material and making it suitable for the computer. Tools like Stylo and AntConc prefer to work with plain ASCII (or UTF-8) text for example. However, the Life Books are always made in MS Word or Adobe PDF format, so the texts must first be converted and cleaned up in a program such as Sublime Text 3, and then saved in flat txt format. Furthermore, the researcher is only concerned with the life story, the narrative, so the title page,

² In September 2019, 25 digital Life Books were received by the Meertens Institute, while the archive now contains 52 paper Life Books; see http://www.meertens.knaw.nl/archieven/index.php?action=expand&querystring_b64=aW50b3VkpWxldmVuc2JvZWsmYW1wO3NIYXJjaF9zdWJtaXR0ZWQ9Wm9law==&id=3918

³ An example of the latter is the ‘digging into data’ project called ISEBEL (Intelligent Search Engine for Belief Legends): <http://www.isebel.eu/site/>

introduction and colophon must be removed. The same applies to photos and irrelevant captions, page numbers, and comments and footnotes from editors. In the end, the 'pure' life story is left over, although this is also somewhat relative: the Life Books never contain completely unprocessed copies of audio recordings, but have always been edited by the interviewers to a greater or lesser extent into logical, fluent stories.

In this pilot I wanted to see to what extent the Life Books can be used for structural analysis, gender differences in narrative style and subject choice, sentiment analysis, recurring themes, distribution of motifs, and perhaps most importantly: thematic gaps. Which (important) issues are *not* raised by the storytellers?

For a (superficial) structural analysis, no computational tools are required. The tables of contents of the Life Books actually immediately show that all the testimonials are extremely chronological. The interviewers and editors may certainly have had a hand in that strict chronology. The story usually begins with one's own parents, then the childhood, the war and its aftermath, school, friends, (for men:) military service, education and occupation, marriage, children, re housings, vacations, illnesses and deaths of loved ones. Yet, some close reading shows that the stories are never the same on a detailed level, despite the emphasis on particular themes: a childhood, for instance, can be happy, carefree, adventurous or filled with misery and mishap. We are dealing with unique human testimonials every time.

At the time of the research I had 19 Life Books available: seven storytellers are men who live in Almere (or lived: some storytellers have since passed away). Twelve storytellers are women: nine from Almere and three from Zaandam. The oldest storyteller is a woman born in 1909, the youngest is also a woman born in 1970. However, most storytellers were born in the 1920s and 30s. Some storytellers were born in small towns or villages such as Olst, Grootegast and Scheveningen. More storytellers were born in big cities, or at least lived a large part of their lives there. This concerns cities such as Amsterdam, Rotterdam, The Hague, Hilversum, Enschede, Nijmegen and Zaandam. All storytellers ended up in Almere or Zaandam, where volunteers from Humanitas visited them (in their homes, homes for the elderly or care centers).

Stylometrics: the style of narrators or interviewers?

With the help of the program Stylo⁴ (which runs in the statistical software R)⁵ I wanted to perform some experimental clusterings to see if there is a stylistic difference between men and women. In this experiment it must be stressed that the interviewers and editors (all amateur volunteers) of Life Books may very well have occurred as a filter: their stylistic characteristics and personalities may play a role in the analysis to some extent. On the other hand, the analysis of language expressions by Stylo can be based on (the unaware use of) function words, and less on content words. In other words: Stylo can mainly look at the use of Dutch words and word combinations such as *de, het, een, op, over, naast, onder* (the, the, a, on, over, near, under) et cetera. If storytellers are quoted literally, emphasized in the layout, these kind of function words will not be lost quickly. Still, in many cases the words of the interviewees are paraphrased, not quoted literally. In my experiment I looked at clustering based on word 2-grams and character 3-grams (combinations of two words

⁴ See <https://journal.r-project.org/archive/2016/RJ-2016-007/RJ-2016-007.pdf>

⁵ See <https://www.r-project.org/>

and combinations of three characters respectively). The result of the clustering was always that there arose separate groups for men and women (sometimes slightly mixed), that the group of women from Zaandam clustered separately, and that two narrators (Sipon and Van Loenen) were systematically regarded as outsiders. Below one clustering as an example:

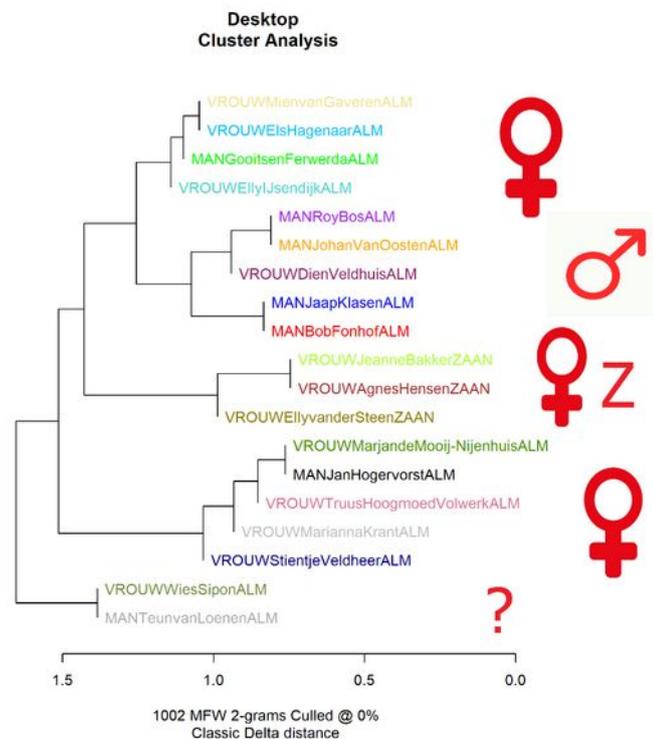


Figure 1: Cluster analysis of 19 life stories based on word 2-grams. From top to bottom we see a predominantly female group, then a group of predominantly men, then the group of women from Zaandam, again a predominantly female group, and finally two outsiders (who stand alone even more distinctly in other visualizations).

All analyses seem to show that men and women have a different style of storytelling, but the gender and personal style of the interviewers must have had some influence on the clustering results as well. In the first ‘female’ cluster the interviewers were Boudewijn Vossen (M), Janneke Wiegiers (F), Jacqueline Streppel (F) and Carla Prins (F). However, in the second ‘male’ cluster we find Carel de Vink (M) as an interviewer *three* times, while the other interviewers for the second cluster were Lex Slager (M) and Jolien van den Heuvel (F) with Willem Jan Hagens (M) – so almost all male interviewers, which may not be a coincidence. Most conclusive is the fourth ‘female’ cluster where Ina van der Vaart (F) was the interviewer and editor in *all five* cases. This test clearly shows that stylistic influence from the interviewers and editors can not be ruled out.

Two other issues still require an explanation. First: why are the women from Zaandam (the third cluster) separate? It is not the case that a regional language from Zaandam seeps through in the texts: all texts, both from Almere and from Zaandam, are written in the Dutch standard language. However, there is one clear stylistic difference. All the Life Books from Almere are written entirely in the first person singular: "My name is X, I was born in the year Y, and I saw the light of day in city A". All Life Books from Zaandam are

(with the exception of a few literal quotes) written in the third person singular: "Mrs. X was born in the year Y, and she saw the light of day in city A".⁶

A second question is why Mrs. Sipon and Mr. Van Loenen always stand apart as storytellers in the visualisations (often even more than in Figure 1). There are no typical similarities: they are of different gender, their interviewers were also of different gender⁷, and they are not exactly the same age (born in 1920 and 1932 respectively). They were both born in Amsterdam, but that applies to more storytellers, who do not cluster with the two. There seems to be only one rational explanation: their life stories are by far the shortest, 3867 and 5693 words respectively. In both cases it concerns Life Books with many photos⁸ and not too much story. Stylo has probably set the texts of Sipon and Van Loenen apart due to a lack of sufficient language material, or clustered them apart as 'short stories'. Stylometric analysis of Life Books is not very helpful in clustering narrative styles. The length of the texts and the influence of the interviewers produce too much noise to analyse the specific narrative styles of the storytellers, for example in connection to gender.

Sentiment analysis

This brings us to the question of who on average tell longer stories about their lives: men or women? It turns out that once men start to talk about their lives, they on average tell more elaborately. Female storytellers have an average of 14,229 words, male storytellers use more than twice that number: an average of 38,139 words. Figure 2 shows that in particular four extensively telling men are responsible for the high average.

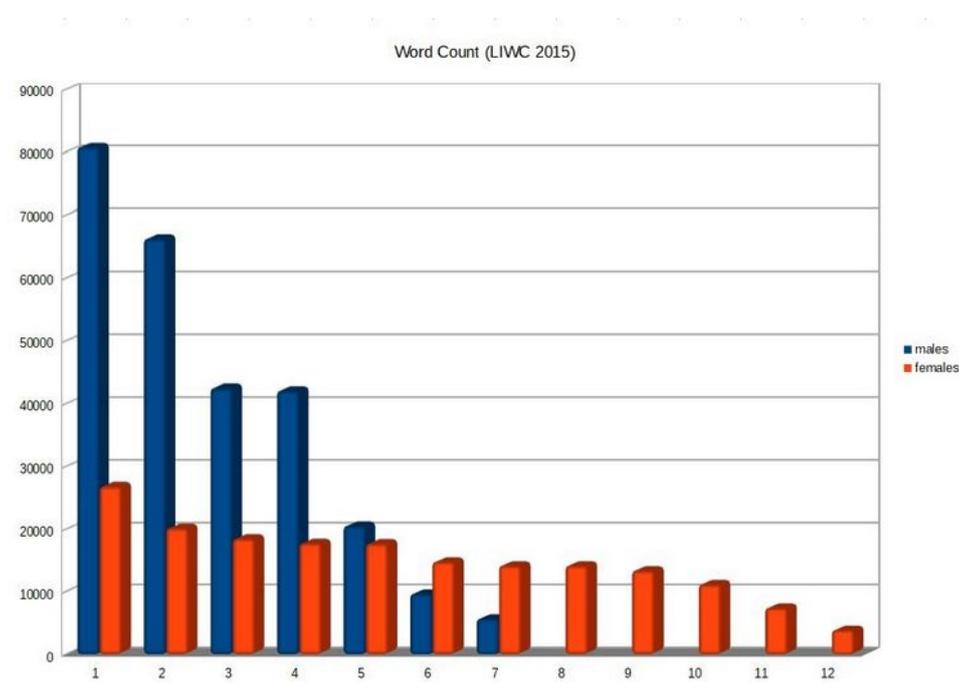


Figure 2: Word Count of the 19 life stories by LIWC2015, visualized in MS Excel. Blue are the men, orange are the women.

⁶ The women from Zaandam were not interviewed by the same person: two were interviewed by Annemieke Blom (F), one was interviewed by Lidwien Berkhout (F). Gender of the interviewers may have played a role though.

⁷ Sipon was interviewed by Carla Prins (F), Van Loenen was interviewed by Willem Jan Hagens (M).

⁸ These photos were deleted for the analysis of the life stories.

With the LIWC2015 program, sentiment analyses can be made.⁹ LIWC2015 operates in the following manner: LIWC has a Dutch lexicon in which all kinds of (mainly emotionally charged) words are classified in certain categories. For example, (a form of) the Dutch word “huilen” (to cry) scores in the following categories: verb (cat. 20), affective emotion (30), negative emotion (32) and sadness (35). All words in a category are added together, then divided by the total number of words in a text and finally multiplied by 100. A score of 0.20 in the 'sadness' category is then relatively low in many cases, while a score of 5.20 is quite high – however, it is particularly important to value the scores by comparing texts. In essence, LIWC is a word counter that is reasonably good at displaying trends in big data. I had LIWC2015 perform analyses on the following categories: positive emotions, negative emotions, anxiety, social processes, family, work, death and sexuality. In the field of emotionality, women always score higher than men, and it does not matter whether they are negative or positive emotions; for women, social manners, emotions, and certainly also fears, are easier to discuss. We see an even clearer difference between men and women when talking about family. Figure 3 shows that women bring up the family much more often.

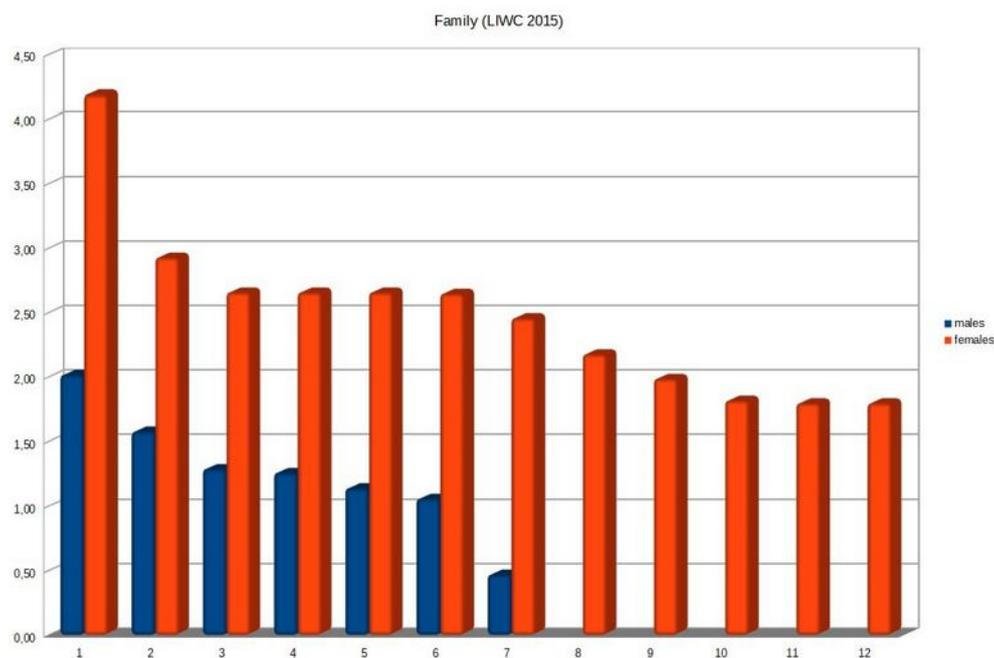


Figure 3: Women and men about “family” in 19 life stories analysed by LIWC2015, and visualized in MS Excel. Orange are the women, blue the men.

One would expect that it would then be the men who have talked about their work a lot, as opposed to women, but in reality the men and the women do not differ that much. After the Second World War, women also started working hard, albeit sometimes in more caring professions as nurse or midwife. The social practice that women had to resign when they got married has gradually fallen out of favor.¹⁰

⁹ See <http://liwc.wpengine.com/>

¹⁰ For such developments, see also Stokvis 1999 (here in particular p. 134); the history that he outlines on the basis of surveys of secularization, lifting of sociocultural barriers, working conditions, upbringing, democratization, leisure, vacations, sex education, divorce and the like largely corresponds to the stories told in the Life Books.

Death is considerably more prominent in the life stories of women than of men. This has the rational explanation that on average women live longer than men: by the time women were interviewed about their lives, their husbands had often already died. But women also tell more about children, brothers or sisters who died young.

Women also tell more about sexuality than men, but here one has to be careful when interpreting the scores that LIWC gives. In this case the values (see the Y-axis in Figure 4) range from 0.02 to 0.15, and these are very low values in comparison with the other outcomes. In fact, it can be said that neither women nor men have touched on the subject of "sexuality."

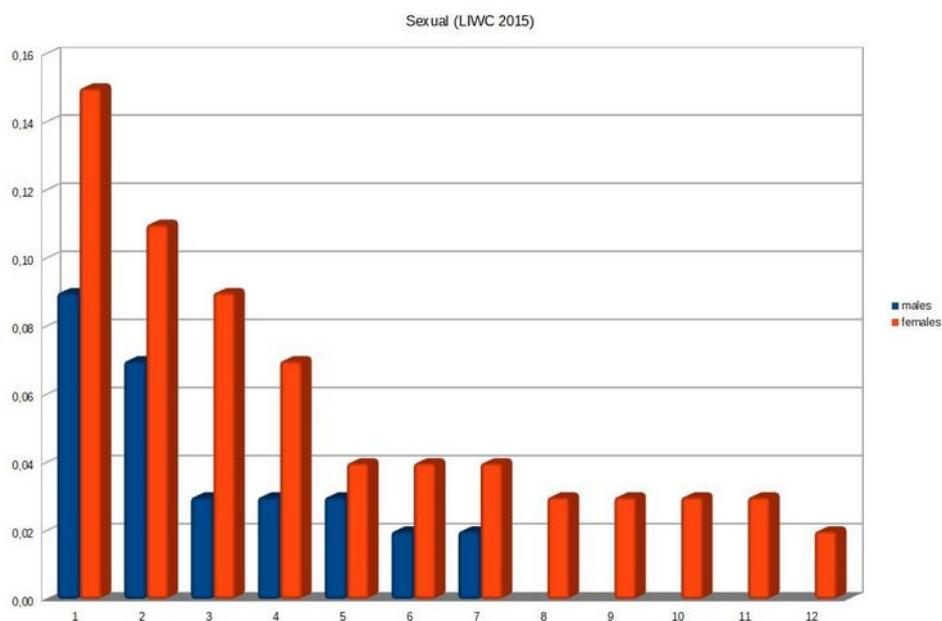


Figure 4:

Women and men about "sexuality" in 19 life stories analysed by LIWC2015, and visualized in MS Excel. Orange are the women, blue the men. Note: the values on the Y-axis are remarkably low here.

Based on the analyses of LIWC, we can state that men on average tell longer life stories than women. However, it is women who use more emotional and affective words in their life stories. Family is a very important topic in the life stories and that is above all emphasized by women.¹¹ When people come to talk about work, the stories of men and women do not differ much from each other. Sexuality appears to be a difficult topic for both men and women to talk about.

Finding sexuality, the war and frequently used words

AntConc¹² is a computational concordance tool to find frequencies and dispersions of words: it can help to locate sections in the stories that otherwise would have been more difficult for the researcher to identify. If one uses the AntConc with wildcards to search for

¹¹ Making a word cloud of the male life stories and the female life stories separately, one sees that words such as "children", "people" and "mother" appear prominently among women. "Father", "mother" and "children" are also clearly visible in the word cloud of men, but dominating are (conjugations of) verbs such as "go", "have", "come" and "must".

¹² See <https://www.laurenceanthony.net/software/antconcl/>

* *sex* * (with which a word as "homosexuality" is found as well), one sees that the concept does not appear in 13 life stories at all. Only three men and three women touch the concept a few times. This does not mean that all sexuality has now been found in the stories, because a search with wildcard for *abuse** also yields 14 hits. It is mainly women who have told about this. One of the women¹³ tells how she was abused as a child (after finding the word, we can now start close reading the text passage):

I was abused between the ages of 5 and 9. [...] To get some extra income, my father let homeless people sleep on our attic floor. They then paid for board and lodging. They also ate here and "played" with me in the attic, or I was "allowed" to sit on their lap. I felt that I was woken up in the middle of the night, but it was probably late in the evening, under the guise of being allowed to come and eat some nice fries, sometimes with chicken. They were certainly not nice men, most of the time they had already been drinking. I was the only one of the children with whom they did that. I wanted to protect my younger sisters. I did not want them to be taken out of bed. I received money from such a man and I handed it to my mother. That way she had the chance to buy food again and my sisters were protected.

In the youth of many storytellers, incidentally, there is still much naivety and ignorance in the field of sexuality. For example, another woman (in the quote: X) tells about the nurse training, where she shared a room internally with another female student:

They also had to call each other nurse. She loved the tour of the three buildings. One building to sleep, one to work and in the third building they ate. Everyone was given a room with several people. X as the youngest got a room in the main building above Obstetrics, together with nurse Y. A nice big room with a bath. Y and X got along really well, they eventually became friends for life. They coordinated their shifts and did homework together. If they were free, they could decide everything themselves. They thought of taking a bath together after work and doing their homework and eating there. That was very pleasant and effective, they thought. But the head nurse, who heard that, thought differently. She called the ladies and interviewed them. Finally she made it clear to them that this was not allowed. X and Y did not understand but did not bathe together any more. They had never heard of homosexuals or lesbians. She found out later when one of the male nurses openly told her that he was gay. Later it turned out that all male nurses were homosexual.

Let us return to an earlier issue: if men have so much to say in their life story, what are they talking about so much? The Second World War. For the boys who survived, it all seems to have been an exciting boy's book in hindsight. For different men, (more than) half of the life story is about the adventures during the war. Figure 5 represents the distribution of the concept * *oorlog* * (war) via the visualization of AntConc (the wildcards also find "wereldoorlog" (world war) and "oorlogsgeweld" (war violence), first with three

¹³ For reasons of privacy, the source of the quote is kept anonymous (also further in this contribution).

men and then with three women. Like a kind of bar code, this visualisation or ‘Concordance Plot’ shows how the concept of war is distributed over the entire life stories:

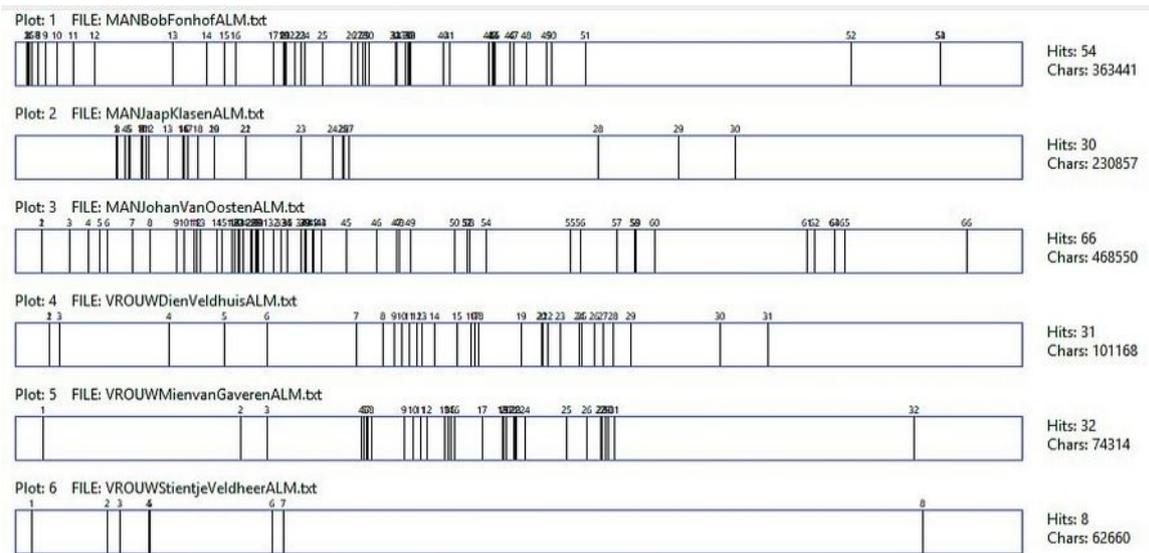


Figure 5: The distribution of the word * oorlog * (war, with wildcards) in the life stories of three men and three women. Visualization with AntConc, mentioning the number of hits, the position in the story, and the number of characters in the entire text.

If we make visualisations of the distribution of the word "ik" (I), the plots are sometimes coloured black for entire blocks. This is of course not surprising when people tell their own life story. In Table 1 I summarize what are prominent content words in the corpus of 19 life stories with regard to themselves, the family, relatives, friends and the immediate environment.¹⁴

Word	Hits
I	12.495
We	4.860
Child(ren)	1.043
Mother	1.018
Father	993
House	898
Family (+ relatives)	483
Friend (m/f)	388
Parents	366
Brother	293
Daughter	240

¹⁴ The function words are omitted, and not all content words are included in this overview, such as "year" (1050 hits), "time" (826), "people" (586) "home" (525) and "school" (470).

Son	152
Sister	125

Table 1: Important content words from the frequency list of AntConc about all 19 life stories that relate to the person, family, relatives, friends and the immediate environment.

Searches in AntConc also show more important moments in the life stories of the storytellers. Rehousings and marriages play a significant role, but also church and vacations. Many storytellers grew up in a Christian environment as a child, and by no means everyone abandoned the church, despite the later development of secularization. In the early days, holidays were mainly cycling holidays in the Netherlands. As prosperity and leisure time increased, and people had access to cars, the holiday destinations became further away. The flying holidays eventually brought people to faraway places, which have left lasting memories.

Memorable events	Number of narrators	Hits
Rehousing	19	207
Marriage	18	204
Church	18	203
Vacation	17	167
Love	17	124
Christmas	12	79
Sinterklaas (Saint Nicolaas)	12	39
Baptism	11	38

Table 2: Memorable events in the corpus of 19 life stories, found with the Concordance Plot function of AntConc.

At least as memorable in the life stories – but in the less positive sense – are the diseases and the death of close ones and loved ones. All 19 speakers talk about illnesses (a total of 558 hits) and all 19 speak about deaths (334 hits). The deaths are mainly at the end of the story when partners die, but also at the beginning of the story, when the parents or very young children die.

Since this is only a pilot study, much of the following results could have been found by close reading and counting by hand as well. However, as soon as the amount of life stories starts growing, computational tools like LIWC and AntConc will be helpful to identify some features of life stories more systematically and more quickly within big corpuses: emotions, important affective relationships, major events in people’s lives, and the impact of war. With the exception of the youngest narrator, who was born in 1970, all the narrators talk about the Second World War – shorter or longer – and the impression this has made on their lives. The oldest storytellers also talk about the years of crisis that preceded it. These are events that also made the history books. But also after the war there were major cultural, economic and political developments that were recorded in the

history books. One somewhat younger narrator remarks that he was a fan of the Beatles, and that he liked to listen to pirate radio station Radio Veronica. AntConc facilitates search for all kinds of details hidden in 'big data'.

Thematic gaps: items not found

However, much of what has been recorded in the history books as memorable events is not to be found at all in the 19 Life Books: AntConc was used for many queries. Where are the "nozems" and the hippies, and where are the Rolling Stones? Where are the experiments with drugs such as weed, speed and LSD? Where's the assassination of President Kennedy, or the first landing on the moon? Where are the "Dolle Mina's" (radical feminists), where are the pill and abortion ("Boss of my own body")? Where is the occupation of the "Maagdenhuis"¹⁵? Demonstrations and protests are nowhere to be found, not even against nuclear energy or the neutron bomb. There seems to have been no oil crisis. There is no sign of famine in Bangladesh. Illustrious Dutch politicians such as Joop den Uyl, Hans Wiegel and Dries van Agt do not feature in the life histories.

Confrontations with events in the world only occur incidentally. At some point an Amsterdam woman says:

I have never noticed anything of all those stories about "Provo"¹⁶ and protests by students. I saw riots at [queen] Beatrix's wedding on TV. I have no memories of the cold war.

Another woman tells about the first house they bought:

A gorgeous house with a bay window, double glazing, stained glass doors, very beautiful. There was a front garden and a large back garden. Squatters had occupied the house, it looked a mess at first. Brown paint had been thrown against the ceiling. [...] The squatters had cats. The house was full of fleas! When X and I entered the house, the fleas jumped against our legs. We deliberately put on white trousers so that we could see them well and catch them. After that, Y disinfected everything. I have vacuumed out all the seams in the wooden floor with the vacuum cleaner. So we finally got rid of the flea plague. The neighborhood was very happy that we bought it. Among others, there was a general practitioner, a veterinarian [...] in the street, and there also was a pharmacy. So those people were only too happy that the squatters were gone.

Here we come across a piece of history from the 1980s, but the squatters' movement is not discussed here because of the protest against the housing shortage, merely because of the negative image of disorderly and destructive living behaviour.

One male narrator mentions the attacks on the Twin Towers on September 11, not in the context of the political and religious clash between the US and Orthodox Muslims, but merely because it ruined his vacation in Los Angeles and San Francisco:

¹⁵ Once an orphanage, later the administrative centre of the University of Amsterdam. Students occupied the building in 1969 to demand more public participation and democracy.

¹⁶ A provocative and anti-authoritarian youth movement (1965-1967).

On 11 September we would fly back – the notorious 11 September 2001, with the attack on the Twin Towers – and that flight was cancelled. Air traffic stopped. When we looked out of our hotel window we saw that the Golden Gate Bridge was occupied by soldiers. On the other side was the state government, there were all soldiers too. X suggested driving back to Los Angeles, because they still flew from there. Along the way we slept with four of us in one room. Arriving in Los Angeles, it turned out that there was no flying from there either. We were able to book a hotel for one night and then we had to leave because it was fully booked. We drove to another hotel and stayed another night there. All in all we were a week later. Fortunately we were able to keep the car and return it to the airport. We were waiting at the airport from nine to half past three before we could finally fly home again.

One storyteller talks about pornography, but he has had a sex shop for a few years. Another briefly mentions the “BOM” mother (Single Parent by choice). One woman speaks briefly about the PvdA (Labour Party) and the VVD (Liberals). And one man mentions the Iron Curtain. These are only minor details in large life stories.

It very much seems that after the Second World War the developments in the world passed by people. We must not forget that they were asked to tell their personal life story. People were not involved in or particularly interested in social developments. Those things happened outside of their "bubble", so to speak. Some phenomena and events were not as big as they appear in the history books – those who were not at the right age, in the right place, and with the right attitude, could easily have missed it. It may also take another generation for the apparently lost stories to be released. Perhaps the events at the time did not go past people, but certain things had lost their importance when narrators reached the age of 75 or 85 years.

However, one narrator is a clear exception: his life story shows that he certainly did not live in a "bubble". This narrator shows to possess political and historical awareness. He entered the army after the Second World War and was sent to Indonesia to participate in so-called “politioele acties” (armed corrective actions) against rebellious militants. The narrator realistically tells about the fierce battle back and forth, and about the mutual atrocities. Gradually he realized that the Dutch government was acting unlawfully, that it should refrain from further colonialist activities, that the insurgents were in fact freedom fighters, and that the Indonesian population had the right to self-determination and a sovereign state. The narrator also speaks expressively about the aftermath of the decolonization process. About the arrival of the KNIL soldiers¹⁷ to the Netherlands, about the immigration of the Moluccans and the promise of the Dutch government to ensure a free Republic of the South Maluku that never came. And he speaks about the train hijacking by Moluccans at De Punt in 1977. That does not mean that his involvement in world events is over. In the 1970s he came to work for the Ministry of Education in Nigeria. He talks elaborately about the geographical, political, religious and social situation there, about the tribal struggle and about the rivalry between Muslims and

¹⁷ Soldiers of the Dutch Indonesian army.

Christians. And finally about the civil war and famine in former Biafra (in southern Nigeria). Such an analytical view of important events in world history is missing in the other 18 Life Books.

Conclusion

Has the pilot with the Life Books been successful? Can computational analyses be made on a much larger collection of digital Life Books in the future? This experiment has shown that it is possible to do research into narrative structures, although this could be much more refined in terms of events. Stylometric analysis with Stylo of male and female repertoires is rather tricky, because interviewers / editors can (very much) interfere as a filter here. Stylo looks for patterns in the use of function words to determine different styles, but Life Books are just not quoting narrators literally, so in quite some cases linguistical features, like the use of function words, may not originate from the storytellers but from the editors.

On the other hand, sentiment analysis in combination with gender, for example, is possible using LIWC2015: this tool can give a fair representation of emotions, relationships and related motifs in life stories. Furthermore, AntConc proves to be a useful tool to investigate the occurrence and distribution of themes and topics. Research into the lack of certain themes and motifs remains an interesting option too. Finally, it can be concluded that the Humanitas Life Books are highly suitable for research into the culture of everyday life, but that there is not yet much detailed information about folktales, songs, games, celebrations and rituals in them.

Biography

Theo Meder – PhD, Senior researcher of Dutch Folktales at the Meertens Instituut in Amsterdam and professor of Dutch Folktales and Narrative Culture at the University of Groningen. His research interests include: folklore, folktales, ethnology, philology, medieval literature and culture, (inter)national databases and computational humanities. Email: Theo.Meder@Meertens.knaw.nl and T.Meder@Rug.nl

Literature

[Collection of Humanitas Life Books](#), archive Meertens Instituut, inv. nr. 496.

Stokvis, Pieter R.D.: *Terugblikken op het huiselijk leven in de twintigste eeuw. Een verzameling getuigenissen over veranderingen in levensstijl sinds 1920*. Leiden 1999.