

Keywords in context: The limits of concordances

Wybo Wiersma

k0931011

King's College London

mail@wybowiersma.net

1 Introduction

In this essay we will be looking at Hannah Arendt's *Eichmann in Jerusalem: A report on the banality of evil*. The book reports on the trial of the German war-criminal Adolf Eichmann. He was in charge of the logistics of the Holocaust and was convicted to death in Jerusalem in 1961. Arendt, the author, was a Jewish political philosopher who fled Germany when Hitler rose to power.

The method we use to examine Arendt's text is concordancing. Concordancing is — in a nutshell — the collating of all occurrences of selected keywords and their immediate context. This is done in order to arrive at precise and/or new interpretations of their usage and ultimately of the text itself. Through this method we will be probing for why *Eichmann in Jerusalem* caused such a controversy, and in the process we will be exploring and delineating the limits of concordances, which is the main purpose of this essay.

First we will go into a bit more detail about Arendt's book and how it could become controversial. Then we will quickly sketch Arendt's view of Eichmann. Following, we will explain concordancing in detail, give our definition of it, and point out the limits of our approach. Then we will discuss some problems with regard to significance that concordancing raises, and introduce a statistical method that can help us select significant keywords. In the next section we will then criticise and partially set aside the usefulness of statistical significance again, and contemplate the nature of the meaning that can be captured by concordances.

Finally we will present our findings on four

levels: that of the process of concordancing, results obtained by concordancing, results provided by statistics, and results obtainable by traditional (philosophical) reading. For each level limits will be given. Then we will wrap up with a discussion of the reasons for our results being rather limited, suggest some possible improvements, and end with a few concluding remarks about some wider implications.

2 Text: Eichmann in Jerusalem

2.1 Controversy

Arendt originally wrote her book as a series of reports for *The New Yorker*. It caused big controversies as it appeared over February and March 1963. Arendt's book was most criticised in America, where she worked at the time, but it also caused quite a stir in Europe and Israel. With one of the more extreme reviews titled: "*Self-hating Jewess Writes Pro-Eichmann Series*" (Dossa 1984). There were three major reasons for the controversy. First of all, Arendt presented Eichmann's perspective and even rationalized it on some occasions. Secondly she was critical of Jewish Councils which had cooperated with the Nazis in organizing deportations: '*if the Jewish people had really been ... leaderless, there would have been chaos and plenty of misery but the total number of victims would hardly have been between four and a half and six million people*'. And finally the main thread that runs throughout the book: Eichmann's evil was not that of a psychotic demon, but that of a small-minded follower, which appeared to excuse him.

The controversy was worsened by the fact

that Arendt interpreted most of the events in terms of her rather unusual political philosophy, without making this very explicit. Critics (and even some supporters) missed her distinction between public and private life, and the different roles that morality and responsibility play in these. Where Arendt considered the cooperation of the Jewish councils supporting Nazi policy: *'politics is not like the nursery; in politics obedience and support are the same'*, she did not mean moral but political support. Similarly, she saw Eichmann's deeds as being induced by his overly strong concern with his own personal and family-interests (his private sphere). Which Arendt explained in terms of the destruction of the public sphere by the totalitarian Nazi regime. She nevertheless held him fully responsible on a moral level.

Arendt had worked out these subtle distinctions in her *Origins of Totalitarianism*, which had appeared in 1951. But few of her critics bothered to read it. Many did not even completely read *Eichmann in Jerusalem* itself. Such confusions and sloppy reading by many of her critics makes that, by limiting our attention to the original text, we will be unable by definition to find the whole cause of the controversy, as much of it likely is in secondary literature that misinterprets Arendt's book.

2.2 *Eichmann*

Eichmann's main task was the scheduling of trains that deported the Jews to and between concentration-camps in Eastern Europe. Many of these camps were death-camps where Jews were killed by gassing on arrival. Besides this, Eichmann had direct authority over only one camp: the Theresienstadt ghetto, a camp on German soil. It was a "model camp" in which high status Jews, and Jews with international connections were incarcerated. It even featured in some Nazi propaganda films.

Arendt argued that Eichmann was not an abnormal man: *'everybody could see that this man was not a "monster" but it was difficult*

indeed not to suspect that he was a clown'. He had been tested by six different psychologists which all were unable to find anything wrong with him. His attitudes to his family and friends were *'not only normal, but most desirable'*. If anything was wrong it was that he was obsessed with his career. Even towards his Jewish interrogators he complained several times about not getting beyond the rank of lieutenant colonel: *'[I] had done everything, even asked to be sent ... "Off to the front, I said to myself, then the Standartenführer [colonelcy] will come quicker"'*.

Moreover, according to Arendt, Eichmann was not a Jew-hater. He had Jewish family, whom he landed his first job at the Vacuum Oil Company. He even had a Jewish mistress when he was working for the S.S. in Vienna in 1940. In addition he had a highly developed bureaucratic conscience. He hated sadists, lies and corruption (the latter became very common in the German army near the end of the war). He saw honour in being efficient and fulfilling his orders to the fullest extent.

Eichmann claimed to have lived according to Kant's Categorical Imperative, which he could reproduce from memory — and roughly correct — as: *'the principle of my will must always be such that it can become the principle of general laws'*. He said he only had been his own law-giver until the "final solution" became policy. Then his Kantianism was reduced to *'Kant "for the household use of the little man"'*, or as Hans Frank formulated it: *'Act in such a way that the Führer, if he knew your action, would approve'*.

Eichmann joined the S.S. in 1934 at the advice of a friend. His first job was the filing of information on Freemasonry, and soon also on Jewish affairs. Then, in 1938, he was moved to the Jewish Emigration office in Vienna. Here he was very successful at navigating bureaucratic obstacles to "help" (force) a hundred and fifty thousand Jews to emigrate to Palestine (Israel), the United States and other destinations. He experienced this period as the zenith of his career, but when, first due

to the war and then due to a ban, emigrations came to a standstill in the summer of 1940, and his plans for evacuating four million Jews to Madagascar was not taken up by 1941, he gave in.

Nevertheless, when Eichmann had to ship his first train to what he knew to be a death-camp, he re-routed it to a Jewish labour ghetto in Lotz instead, causing him trouble (Lotz was already overcrowded). From then on all his trains ran with the greatest efficiency. Even when Himmler, in defiance of Hitler, decreed a stop to deportations in order to be able to destroy evidence of the Holocaust, and to attain international goodwill, Eichmann continued shipping Jews. So while Arendt declared Eichmann's evil to be banal, the banal evil had demonic consequences nonetheless. The death of millions of Jews cried for a criminal that would fit the crime. Much of the reaction to Arendt being rooted in (perfectly justified) trauma and emotion and makes it go beyond language, and thus beyond Arendt's language as well.

3 Concordancing

3.1 What

Concordances are listings of occurrences of keywords with their immediate context. They are produced from the source-text by a computer program. There are several types of concordances. First of all there is a difference between semantic and linguistic concordances. The former, in which no keywords but key-concepts are used, can not be produced automatically by computers. So they only have been crafted for canonical texts such as the Bible. Secondly, concordances can come with, or without, context-words from the text. Linguistic concordances with context-words are called KWIC (KeyWord In Context) concordances. They are the focus of this paper. Finally, the context can consist of the whole sentence, or of a span with a fixed size: in words or characters. We used a span of 73 characters on each side of the keyword to make optimum use of the screen-width (of

160 characters).

When making a KWIC concordance (henceforth called concordance) all occurrences of the keyword in the text are found, and presented with their immediate context. The aim of this is to restructure the text so all occurrences of the keyword can be seen at a glance, providing a new perspective, leading to new and better interpretations. An example of some concordances from Arendt's text can be seen below (each spread over three lines):

```
unsuccessful effort to prove that
    Eichmann
        had once , at least , killed
am made out to be",
    Eichmann
        said "I am the victim
nothing but a regrettable mistake ,
    Eichmann
        was troubled by no questions
```

By only considering KWIC concordances we thus define concordances narrowly. Searching through the text, occurrence by occurrence, using one's word-processor, as well as things such as Google searches, are not considered concordances here.

3.2 Tool

While there are many good concordancing tools available, both commercially and for free, such as: *Concordance*, *MonoConc*, *WordSmith* (commercial), *AntConc*, *Xaira*, and *ddc-concordance* (free), a concordancing tool was created from scratch for this research. First of all because a custom made script would be easier to adapt. Secondly, because it was a nice challenge, and given the availability of modern third generation programming languages such as Python, it was not much work (about 15 hours). The latest version of the script, *conq.py*, is only 140 lines long (included as appendix A). Thirdly, a script that works under Linux and that is console-based was preferred, and those are rarer.

Regardless of the commonly professed advantages of Graphical User Interfaces (GUIs), the console remains a versatile and fast environment for text-processing. It offers a host

of text-related tools such as `wc` (word-counts), `sed` (search-replace), `less` (viewing) and `vim` (editing), which can be made to work together easily through `|` piping, and `>>` redirecting (sending output of one program directly into another, resp. into a file). In addition, when one will be using the same tool over and over again, the learning-curve of the console becomes worth it. At the very least, not having to use the mouse all the time reduces the risk of RSI.

The `conq.py` program has all the basic features of a concordance-tool: variable context-spans, regular expression search, multi-keyword search, key-phrase search, and sorting: by keyword, and by left and right context. In addition it automatically logs and saves all the searches. Thus making it easy to keep a logbook of searches and the results they delivered. Finally, thanks to the speed of current-day computers it is fast enough: it scans the hundred and ten thousand words of Arendts text in about two seconds.

In this research we limited ourselves to what could be produced using this simple tool (apart from the statistical method introduced next). Advanced features such as lemmatizing (grouping words with the same root together), or UI-related features such as hypertext-links between words in the text and their concordances were not considered.

4 Statistics

4.1 Problems

While concordances can quickly be produced using modern computers, the keywords for which they are generated are normally selected by the scholar. In most cases keywords are glanced from a word-frequency list. A word frequency list is a two-column list showing all words and their frequency in the text, ordered from frequent to rare. The top ten for Arendts text (excluding punctuation) is shown in table 1.

One problem with this method, and similar methods such as using intuition or drawing inspiration from earlier concordances, is that

Table 1: *Top-10 word frequencies and 3 additional open class words*

Frequency	Word
8104	the
4443	of
3279	to
2734	in
2723	and
1999	was
1832	a
1668	that
1569	had
1130	he
...	...
671	Jews
...	...
543	Eichmann
...	...
440	Jewish

there is no guarantee of significance. How to know that the keyword is actually typical for the author or the text, and not just common in general? Another problem is how to go beyond merely finding what one is looking for? E.g. ensuring that one is telling something about the text, and not merely confirming ones preconception of it.

Comparison seems to offer a way out of both of these problems. In it's simplest form one could compare word-counts for Arendts text to word-counts for other texts. A variety of other texts might be used for this. A text that represents average (British) language usage is the British International Corpus of English (ICE) corpus. Comparing in this simple manner is better than not comparing at all, but if ones text has 'his' in place 18, and the other text has it in place 20, how would you know that the difference is not caused by mere chance?

If one wants to be sure of significance it is better to use a statistical method. Besides statistical significance, statistics promises the additional benefit of finding less frequent, but typical words, and thus to present unusual things about the text that one would normally have missed. A suitable statistical method is permutation statistics. We will discuss it next.

4.2 *Permutation Statistics*

Permutation statistics is a relatively unknown branch of statistics invented by R.A. Fisher and E.J.G. Pitman in the 1930s. The method only recently became practically applicable with the advent of powerful computers. W. Wiersma, J. Nerbonne and T. Louttamus (forthcoming) have recently adapted it for finding significant differences in syntax between the English of learners and that of native speakers.

The basic idea of permutation testing is very simple when applied our data: First get word-frequencies (counts per word-type) for each text in the corpus you are examining. Then shuffle the texts of the corpus together with a comparable corpus and draw texts out of the mix until you have a corpus of the same size as your original. Then get word-frequencies again. Do this 10.000 times and track for each word how often it was more frequent in the permutations than in the original corpus. If a word is only rarely more frequent in later permutations, then clearly it was very typical for the original text.

In our example the original corpus was Arendts text, split up in chunks of roughly 2250 words (respecting paragraph boundaries). The size of the chunks is similar to those in the ICE-corpus, of which only the printed non-academic, reportage and creative texts were selected (166.000 words in total).

The “typicality” that the method delivers can even be given a number (times more frequent divided by 10.000) which is equal to a p -value in traditional statistics. As p -values under 0.05 are normally considered significant, we use this as a cut-off value. Thus we get a list of words of which we are reasonably sure that they are typical of Arendt. The first 10 words, when sorted by relative frequency in Arendts text are shown in table 2 (excluding punctuation and names).

They are all words we would expect to find. So the method works. As can be seen in table 3 it also works for collocates (excluding punctuation, numbers, names and closed class

words).

Table 2: *Top 10 typical words*

Word
bureaucracy
Zionists
implementation
decree
presiding
deportations
examiner
offense
evacuation
tribunal

Table 3: *Top 10 typical collocates*

Collocate
death camps
Jews had
Police Leaders
Third Reich
Higher S.S.
concentration camps
General Government
crime against
Jewish functionaries
Jews were

It should be noted that in order to successfully apply the method, some special normalizations have to be applied, which are beyond the scope of this essay, but are explained in the Wiersma et. al. paper.

One remaining pressing problem however is that this method, and any similar statistical method, only finds words that are significant. Thus the concordance-contexts, or any other form of meaning beyond that of individual keywords, are not captured by it. Which makes the usage of statistics problematic: especially the wielding of claims of significance.

5 **Meaning**

5.1 *Text*

The issue of meaning points to another problem with concordances: where, at what level, should we envisage meaning to reside? Is it found in individual words, the concordance

context, sentences, the page, the text, the intertextual, the reader, or even Platos eternal forms? Only if all of it is found in words will claims of statistical significance be valid, but similarly, concordances can only be fully trusted if meaning arises within the concordances span. Only then can we safely claim to understand the usage of keywords from the word- or character-span.

If meaning, on the other hand, arises at the level of the text, and assuming concordancing restructures the text to create a new view of it, we land ourselves at the *Ship of Theseus* paradox. It goes as follows: *'The ship ... was preserved by the Athenians ... they took away the old planks as they decayed, putting in new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers ... one side holding that the ship remained the same, and the other contending that it was not the same.'* Then Hobbes asked an additional question: what happens if we build a modern ship from the old planks? Concordances do exactly this Hobbesian thing if the planks are seen as the concordance-snippets, and the text as the ship(s). Are we still speaking of the same text if our findings come from a different text? Can we safely deduce the ships design by looking at each of the torn out planks in turn?

It is perhaps no wonder that concordances were first made of the Bible, a collection of texts in which every word is considered to be passed down through a divine influence, and thus of the utmost value. In addition, from the fact that paraphrasing a work of Shakespeare, and then deducing something from that new edition, will not even do in worldly literary scholarship, it seems that there is great attachment to the "planks" among literary scholars. In analytical philosophy, on the other hand, it is common practice to reformulate another's text and then to reason from that. Philosophers seem more drawn to the abstract form of the "ship". Concordances will thus be considered less useful by them.

Nevertheless the most likely place where

meaning arises in concordances might be the inter-textual, the concordance in relation to the original text and perhaps other texts, as mediated by the reader. In modern times computer-generated concordances are generally not seen as finished, original research. Intelligent interpretation comes from the scholar bringing his own context to both the concordances and the text, forming a hermeneutic triangle. Meaning arrived at through concordances must thus be of an associative, intertextual kind.

5.2 Computers

Computers cannot handle meaning very well. They can only deal with structure. The best illustration of this is the near start of a nuclear Holocaust on the 5th of October 1960, when a just installed early warning system indicated that a barrage of Soviet missiles was on its way to the United States ...while it was only the moon rising from the east. This example suggests that the problem of meaning in computers is a lack of context: bringing common knowledge about the world into the equation.

Computers are quick banal idiots, or in Peter Wegners terms: algorithms are autistic. It is not for no reason that they flourish in bureaucratic settings. They are the ideal office tools: they are little, fast, Eichmanns. For them importance equals frequency, and efficiency is heir benchmark. They require that what they are dealing with is stable and unambiguous. The subject, whether a keyword in a statistical calculation, or a Jew on a train, is reduced to a number, and that has to be all there is: no word-span, let alone a whole text or inter-textuality around it, which drags the whole world along.

This contextlessness of computers makes that, at least with the current state of AI, we cannot get statistical significance at a level that is meaningful for the literary scholar. Statistics are thus not very useful in a literary context, and if used at all, they are, as McCarty already noted, best used for exploratory purposes; e.g. for being presented with the unexpected. Computers offer new angles and

restructured texts, not certainty, nor the fast processing of meaning. Outside of the office they can only help us to find the unusual.

6 Findings

6.1 Process

While the console offers many benefits, and still is the preferred environment, some of its downsides were experienced during the research. The biggest problem being the two-second time-delay penalty for doing new concordances (a bit more if one counts the time needed for typing in the keyword). As D.F. Galetta has shown (for websites at least), any delay over two seconds is likely to inhibit doing more searches, and thus more creative searches as well. Which makes it more likely that one will only find what one is looking for. Secondly, as the console offers no hyper-text or click-through possibilities, there was a divide between the concordances and the text. Which holds one back from exploring the wider context of keywords.

As noted earlier, a custom tool was developed during this research. Its development was occasionally found to be a distraction from the research. It was quite easy to get drawn into the programming. Especially as its rewards were surer, and more instant.

Some things that would have been useful for concordancing also appeared to fundamentally hard to do. Dealing with homonyms such as “all right” vs “morally right” vs “legal right” is one example. Being able to exclude concordances containing specific context-words (say “legal” when searching for “right” to get at the morally “right”) could have alleviated this (and is not impossible to code). Another, more fundamental problem, is making sure one has added all synonyms to the multi-keyword search command.

Finally, concordancing seems to be one of those things that get better over time. Therefore experience with other texts would have been immensely useful. Given the limited time available to analyse the concordances of other texts, this concordancing project was

limited by some degree of inexperience.

6.2 Statistics

Applying permutation statistics to Arendts text resulted in a list of 586 typical words and 748 typical collocates (typical relative to the ICE Corpus). We have already listed some of them in section 4.2. At first sight this seems to be a great result. But as the words were inspected it appeared that most of them are names and places, and words that set the topic apart such as “Jews”, “deportations” and “offense”. In other words they were mostly as expected.

The first of the few unexpected things that turned up were related to punctuation. Arendt uses a lot of punctuation. Significantly more than the ICE-corpus. Most typical, in order, are quotation marks ‘’’, semicolons ‘;’, dashes ‘-’ and commas ‘,’. The quotation-marks can be explained by the quoting of Nazi “newspeak”, and contested words, such as: “Gottgläubiger” (S.S. term for atheist), “evacuation” (deportation), “final solution” (holocaust), “language rule” (this “newspeak”). Then semicolons, dashes and commas can be explained by her long sentences (41 words long versus 22 for the ICE texts, both counts including punctuation). Which in turn gives away her German academic background (where long sentences are considered a sign of sophistication).

The only really meaningful surprises are her over-use of collocates such as “to be sure”, “with respect to”, “was indeed”, which Arendt often uses to assert accepted opinion, after which she takes a different angle on it, such as in *‘To be sure, those who resisted were a minority but under the circumstances “the miracle was,” as one of them pointed out, “that this minority existed”’*. Another surprise is Arendts use of the word “fantastic” (both as in “stuff of fantasy” and “great”), which is not used in any of the printed texts in the ICE corpus at all. While these results point to the taking of new angles and to an informal use of language, on overall the statistical results are not “fantastic”.

6.3 Concordances

The results for traditional concordancing were slightly better. From the concordances of the keyword Eichmann, it could be glanced that Arendt often portrayed him as not very bright (let alone cunning or evil). Words such as “confused”, “muddled”, and “inkling” often described his thinking. Then when running concordances on these words it appeared that, surprisingly, Arendt also used them in relation to others (Jews, the prosecutor, and the judges at the Neurenberg trial), as well as to disambiguate things such as the complicated structure of the Nazi bureaucracy (such as “*not to be confused with Himmler’s Security Service*”).

Then for the word “banal”, which was not among those statistically typical for Arendt: It is only used three times in the whole text, and only once (twice if the subtitle is counted) if we exclude the afterword (which was added in the 1964 edition). While obviously it expresses the essence of her book. This is a case in which concordances don’t represent the text well. Another interesting observation is that Arendt in her afterword explicitly criticizes notions of collective responsibility, while she often speaks in terms of groups herself, even in contexts of responsibility: “*Nazi Germany was responsible*”. This could be mere metonymy of course, but it still is there.

Finally, Arendt sometimes uses unusual language. The word “fantastic” was already mentioned, but other out of place words such as “clown”, “funny” and “joke” were found more than five times each, often in relation to Eichmann, sometimes even in relation to the holocaust as a whole: “*calling the whole thing [Holocaust] off as though it had been a joke*”. It is likely that this ironic language (whether consciously or unconsciously) tripped up at least some readers, and thus was responsible for some of the controversy around her book. Nevertheless, given the explicit nature of these words, they can be (and were) spotted by normal reading as well.

6.4 Philosophy

By classical, manual reading of Arendt’s text, a wealth of ideas can be generated. One being that according to Arendt Eichmann was entangled in language: ‘*Eichmann’s great susceptibility to stock phrases ... made him ... an ideal subject for “language rules”*’ and ‘*it amounted to a mild case of aphasia - he apologized, saying, “Officialese [Amtssprache] is my only language”*’. To such an extent even, that, as already noted, he still complained about the limited progress of his career in terms of S.S. morality. Eichmann, just like a limited word-span concordance, was unable to look beyond his context: the functional aspects of his job. He missed the wider moral meaning of his deeds.

This possibility of being locked up in language is something which closely matches the philosophy of Michael Foucault, who claimed that power did not reside in individuals, but in language. In so called discourses, which provide a context for true and false, and implicitly determine what can be said and thought. Foucault even argued that the existence of the individual author is an illusion. Literary analysis seems to confirm the influence of discourses, in that differences in style between genres for the same author are bigger than those between authors.

The sociologist Niklas Luhmann takes the notion of discourses to a next level, by arguing that not people are the driving force behind communications, but communications themselves. People are merely a substratum from which autopoietic (self-reproducing) social systems emerge, which then procreate themselves according to their own internal logic. This includes exclusion-mechanisms, and self-reinforcing tendencies, such as, for example, can be seen in the market: managers who care more about truth or goodness are competed away by those that focus on profit, and only the successful managers get to pass on their values in public speeches. Luhmann claims that society has become steerless, even for those at the helm. People that opposed

were simply excluded.

Eichmann stated that he never had heard anybody say anything against the “final solution” during the war. And given his private and professional environments, this is a credible claim. Who would Eichmann be to question his superiors, many of whom were from more established backgrounds than he? Thus we are wrong in assuming that the criminal always has to fit the crime. The Eichmann trial inspired the Milgram experiments after all, in which a majority of those partaking were persuaded to kill other test-subjects by electrocution. Was it really just to convict Eichmann for not being exceptional? Or was it an exceptional situation in which the hanging of a compromised scapegoat was justified to ensure a comforting distance between us and “the German monster”. Are we, thinking of Luhmann’s view of communications that communicate, like keywords in context? Limited to drawing our meaning from these communications?

To what extent is to understand, really to condone? Is to understand and explain things in terms of the social system not rather than to pardon the individual, to condemn the context? Should we not try to learn to live with the uneasy fact that it could have been us. Given the “right” wrong text and a narrow word-span? Instead of taking the easy route of presuming a diabolical essence in “Eichmann”, is it not better to critically examine, and to improve the moral nature of our own social systems, rather than to change nothing and to only criticise individuals, relegating morality to the margins of society (and to something that demands and induces behaviour that will only marginalize those who stick to morality?).

Given the right inter-textuality we might actually already be behaving nearly as badly as Eichmann. If solely tying responsibility to individuals still will be common practice in the future, it makes me wonder what they will think of us when they take “concordances” of our lives (or the lives of the “Eichmanns” currently in charge of scheduling them), fo-

cus on key concepts such as global warming, the waste of limited resources, and world poverty. This is something that may not be easy to stomach, but most of our cramps here come from considering it would condemn us personally, instead of our context.

Of course such questions are big. Certainly they go beyond the span of concordances. Clearly we should not judge concordances by their inability to throw light on questions so far reaching that they are almost meaningless, even for the span of human minds.

7 Discussion

While we did get some results concordancing Arendts text, they are not very staggering. Several reasons might be found for this. First of all, expectations might have been too high. The novelty of results is, after all, in the eye of the beholder. Given that the text was not very long (only 110.000 words) and it was read twice, even before the concordancing began (plus some background-literature), much of what could be teased out with concordances was already known or expected. Concordancing a longer text, or an unfamiliar text, might have been experienced as being more fruitful. Also, even if concordancing is not better than reading a text, it could still allow one to explore texts with greater efficiency.

Secondly, in addition to getting better with experience, concordances (as well as statistics) should work much better for comparisons between texts than for looking at single texts. In hindsight, comparing *Eichmann in Jerusalem* to, for example, Arendts *The Origins of Totalitarianism*, might have delivered more telling results, both from concordances and permutation statistics. For the statistics it would also have been interesting to juxtapose the text to one on the same topic. In that way it would set Arendts language apart and not the topic. Also comparing the 1964 version of the text to the text as it appeared in *The New Yorker* might have delivered interesting results. This because Arendt allegedly removed some of the strong language used in the early edition. Sadly this was not possible

as the archives of *The New Yorker* are behind a paywall to which the library of King's College London has no subscription.

Thirdly, it is likely that Arendts text itself was not very suitable for concordancing because it was written in a very journalistic style. Both a more classical philosophical text and a more literary text might give better results. And in case of a philosophical text, both one more typical of the continental, resp. analytical tradition would offer more suitable keywords. The former providing neologisms (such as Heideggers "Dasein" or Sartres "angst") to compare for senses and uses across contexts, while the latter would offer more strict and patterned language. Thomas Kuhns work, for example, with his varied use of the word "paradigm" has been a concordancing gold-mine. Arendts text seems to strangely sit in the middle; like the eye in a tornado of controversy.

Fourthly the traditional Platonic dichotomy between truth and beauty might make concordances less central to philosophical analysis than to literary analysis. If one sets out to find things that throw new light on the meaning of the text, instead of things that could clarify how the established meaning has been achieved, one might not find much. For example Arendts usage of "funny" words may be relevant from a literary perspective, and be crucial to how the text was received or even (mis-)interpreted, but it is less so from a traditional philosophical perspective. While limited however, concordances can still offer us something new and unique. Even if just by drawing our attention to commonly ignored aspects of a text.

8 Conclusion

We started out by noting that many of Arendts critics had not read her work thoroughly on one hand, and that, on the other hand, the reception of her text was largely determined by (justified) factors that have nothing to do with the wording of the text itself. Then we defined concordances as KWIC concordances, and excluded other forms of searching. In our

research we also limited ourselves to using a console-based tool written from scratch.

Then we discussed the problem of picking significant keywords, and tried out permutation statistics as a remedy. Following, the limited suitability of statistical approaches for situations in which meaning reaches beyond keywords was noted. Next this problem was extended to concordances in relation to meaning at levels beyond the concordances span. It was concluded that the insights gained using concordances are of a hermeneutic and associative kind, limiting the applicability of statistical significance. This confirmed that computers are most useful as exploratory tools, not for strapping down meaning.

Then the results were discussed. First of all, the console was found not to be without its influence, nor to be ideal (thought still preferred). The statistical approach was found to deliver mostly expected results, but was nevertheless able to tease out Arendts usage of the highly informal "fantastic". Normal concordancing delivered some results, notably showing the usage of inappropriate ironic language such as "clown" and "joke". But most of these latter results would have come to light through careful manual reading as well. Which leaves the main advantages of concordancing at being faster than reading, and focusing ones attention at aspects of the text that are easily ignored.

Finally, a lot has come to light about how concordances could have been used better: From doing comparisons, to applying permutation statistics to different a set of texts. Surprisingly, concordances also offered an useful metaphor for the Eichmann case, and for issues of moral blindness and responsibility in general. To conclude: Limited results, and negative results, are results also. Whether ours, those of others, and whether literal or metaphorical, we can learn from our mistakes, so we will not repeat them. Besides critically probing into the limits of our contexts, all we can do is hope that we may be saved from times of trial, and be delivered from evil, whether intentional or banal.

References

- aConCorde* @ School of Computing, University of Leeds - Andrew Roberts - Software. 2010. URL: <http://www.andy-roberts.net/software/aConCorde/index.html>.
- AntConc* - Laurence Anthony. 2010. URL: http://www.antlab.sci.waseda.ac.jp/antconc_index.html.
- Concordance: software for concordancing and text analysis*. 2010. URL: <http://www.concordancesoftware.co.uk/>.
- ddc-concordance - a search engine for linguists*. 2010. URL: <http://www.ddc-concordance.org/>.
- Diner, D. and R. Bashaw. "Hannah Arendt Reconsidered: On the Banal and the Evil in Her Holocaust Narrative". In: *New German Critique* (1997), pp. 177–190.
- Dossa, S. "Hannah Arendt on Eichmann: The Public, the Private and Evil". In: *The Review of Politics* 46.2 (1984), pp. 163–182.
- Fisher, R. A. "The design of experiment". In: *New York: Hafner* (1935).
- Galletta, D. F et al. "Web site delays: How tolerant are users". In: *Journal of the Association for Information Systems* 5.1 (2004), pp. 1–28.
- Ginzburg, C. "Clues: roots of an evidential paradigm". In: *Clues, Myths, and the Historical Method* (1989), pp. 96–125.
- Good, P. "Extensions of the concept of exchangeability and their applications". In: *Journal of Modern Applied Statistical Methods* 1 (2002), pp. 243–247.
- Hannah, A. *The origins of totalitarianism*. Harcourt Brace, 1979.
- Hirst, G. "Context as a spurious concept". In: *Proceedings, Conference on Intelligent Processing and Computational Linguistics* (2010), pp. 273–287.
- Hockey, S. *Electronic texts in the humanities: principles and practice*. Oxford University Press, USA, 2000.
- Hoover, D. "The end of the irrelevant text: Electronic texts, linguistics, and literary theory". In: *Digital Humanities Quarterly* 1 (2007).
- Kenny, A. J.P. *The computation of style: An introduction to statistics for students of literature and humanities*. Pergamon, 1982.
- Manning, C. D and H. Schütze. *Foundations of statistical natural language processing*. MIT Press, 2000.
- McCarty, W. "Being Reborn: The Humanities, Computing and Styles of Scientific Reasoning". In: *Renaissance Society of America 2005 conference, April 7* (2005).
- "Beyond the Word: Modelling Literary Context". In: *Text Technology* (1998).
- MonoConc and ParaConc - Concordance software*. 2010. URL: <http://www.athel.com/mono.html>.
- Pitman, E. J. G. "Significance tests which may be applied to samples from any populations III. The Analysis of Variance Test". In: *Biometrika* 29.3-4 (1938), p. 322.
- Plutarch. *Theseus by Plutarch - The Internet Classics Archive*. 2010. URL: <http://classics.mit.edu/Plutarch/theseus.html>.
- Rockwell, G. "What is text analysis, really?" In: *Literary and linguistic computing* 18.2 (2003), p. 209.
- Sinclair, J. *Corpus, concordance, collocation*. Oxford University Press, 1991.
- Smith, B. C. "The limits of correctness". In: *ACM SIGCAS Computers and Society* 14.1 (1985), pp. 18–26.
- Wiersma, Wybo, John Nerbonne, and Timo Lauttamus. "Automatically Extracting Typical Syntactic Differences from Corpora". In: *Literary and linguistic computing* forthcoming (2010).
- Wilks, Y. *Grammar, meaning and the machine analysis of language*. Routledge/Thoemms Press, 1972.
- Wittig, S. "The computer and the concept of text". In: *Computers and the Humanities* 11.4 (1977), pp. 211–215.

Wolfson, M. "Constraint and Choice in the SS Leadership". In: *Political Research Quarterly* 18.3 (1965), p. 551.

WordSmith Tools. 2010. URL: <http://www.lexically.net/wordsmith/>.

Xaira, Oxford University Computing Services. 2010. URL: <http://www.oucs.ox.ac.uk/rts/xaira/>.

Appendix A

conq.py:

```
#!/usr/bin/python
import re, string, datetime
from read_words import *

c_dir = '/home/wybo/projects/fiauimenre/corpusData/arendt/wordrow/*.txt'
terms = "moreover" # search-term, "term1, term2, collocation"
sort = "right" # left, right, term of False
reverse_sort = False # reverse the sort order
wing = 20 # maximum number of words on each side
screen_width = 160 # total width of your console-window
default_screen_middle = 14 # amount of space around output

term_sets = []
terms_arr = re.split(",\s+", terms)
for term in terms_arr:
    term_sets.append(term.split(" "))

def execute(c_dir, term_sets, sort=False, reverse_sort=False):
    (words, files) = get_words(c_dir)
    concordances = find_concordances(words, term_sets, files)
    sorted_concordances = sort_concordances(concordances, sort, reverse_sort)
    output = format_concordances(sorted_concordances)
    print "".join(output)
    write_file('logs/' +
               datetime.datetime.today().strftime('%Y%m%d-%H%M%S-%f') +
               term_sets_string() + '-' + str(sort) + '_sorted.txt', output)

def write_file(file_name, lines):
    '''Writes the lines to the file_name'''
    outfile = file(file_name, "w")
    outfile.writelines(lines)
    outfile.close()

def find_concordances(words, term_sets, files):
    concordances = []
    for set in term_sets:
        for w_index in range(len(words)):
            set_len = len(set)
            w_offset = 0
            w_punc_offset = 0
            for term in set:
                while w_offset > 0 and (w_index + w_offset + 1) < len(words) and \
                    re.match("[ " + string.punctuation + "]", words[w_index + w_offset]):
                    # detects compound-words containing punctuation
                    w_offset += 1
                    w_punc_offset += 1
                if re.match('^' + term + '$', words[w_index + w_offset], re.I):
                    w_offset += 1
            else:
                break
```

```

w_offset -= w_punc_offset
if w_offset == set_len:
    if w_index > wing:
        conc = [words[(w_index - wing):w_index]]
    else:
        conc = [words[:w_index]]
    w_end = w_index + w_offset
    conc.append(words[w_index:w_end])
    if w_end - 1 < (len(words) - wing):
        conc.append(words[(w_end):(w_end + wing)])
    else:
        conc.append(words[(w_end):])
    file_pre = files[w_index].split(".")[0]
    conc.append([int(file_pre)])
    concordances.append(conc)
return concordances

def sort_concordances(concordances, sort, reverse):
    concordances.sort(CompareConcordances(sort), reverse=reverse)
    return concordances

class CompareConcordances:
    def __init__(self, sort):
        if sort == "left":
            self.index = 0
            self.reverse = True
        elif sort == "right":
            self.index = 2
            self.reverse = False
        elif sort == "term":
            self.index = 1
            self.reverse = False
        elif sort == False:
            self.index = 3
            self.reverse = False
        else:
            raise "Wrong sort argument"

    def __call__(self, conc_x, conc_y):
        conc_x_c = list(conc_x[self.index])
        conc_y_c = list(conc_y[self.index])
        if self.reverse:
            conc_x_c.reverse()
            conc_y_c.reverse()
        conc_x_c = self.normalize(conc_x_c)
        conc_y_c = self.normalize(conc_y_c)
        if conc_x_c > conc_y_c:
            return 1
        elif conc_x_c == conc_y_c:
            return 0
        else:
            return -1

    def normalize(self, list):
        new_list = []

```

```

for w in list:
    if isinstance(w, basestring):
        w = w.lower()
        if not re.match("[ " + string.punctuation + "]", w):
            new_list.append(w)
        else:
            new_list.append(w)
return new_list

def term_sets_string():
    term_line = ""
    for set in term_sets:
        term_line += "-" + ".join(set)
    term_line = re.sub(r"[\w-]", ",", term_line)
    return term_line

def format_concordances(concordances):
    output = []
    for conc in concordances:
        term_set = ' '.join(conc[1])
        left_string = ' '.join(conc[0])
        screen_middle = max(len(term_set) + 4, default_screen_middle)
        screen_side = (screen_width - screen_middle) / 2
        if len(left_string) > screen_side:
            left_string = left_string[len(left_string) - screen_side:]
        out_line = ("{:>" + str(screen_side) + "s").format(left_string)
        out_line += ("{:^" + str(screen_middle) + "s").format(term_set)
        out_line += ("{:<" + str(screen_side) + "s").format(
            (' '.join(conc[2]))[:screen_side])
        out_line += " " + str(conc[3][0])
        output.append(out_line + "\n")
    return output

execute(c_dir, term_sets, sort, reverse_sort)

```