Digital Humanities – Final project

Magdaléna Musichina, University of Helsinki, winter 2018

Bob Dylan x Arthur Rimbaud

In this project I have worked with two corpora – a collection of poems by the French poet Arthur Rimbaud and a collection of lyrics by the famous American musician Bob Dylan. During my studies of literature, I have stumbled upon this unexpected connection a couple of times and thought that this project might be a great way of finding out more about it. Arthur Rimbaud was a French poet in the 2nd half of 19th century. He is said to be an „Accursed poet" just like Charles Baudelaire and Paul Verlaine, led by the principles of symbolism.

Bob Dylan on the other hand is a contemporary American rock (but also folk, blues, country,…) singer and song-writer. His songs include many meaningful themes, ranging from political songs to love songs. He has been awarded the Nobel Prize for Literature in 2016 and in an interview regarding this he listed Rimbaud as one of his most important literary influences. Dylan also stated before that Rimbaud was his favourite poet. Reading Rimbaud's poems and listening to Dylan's music, one can notice they both share a similar spirit of rebellion and grandeur, but there is also the underlying sense of sadness and melancholy. However, can this very abstract feel of the songs and poems be picked up and confirmed by text analysis? What themes do they really share?

Data used

The Rimbaud corpus was formed using a collection of his poems which featured their English translations. The Collected Poems by Arthur Rimbaud has been published by the Oxford University Press and it is accessible online (at https://epdf.tips/collected-poems-oxford-worlds-classics.html).

The Dylan corpus was a bit harder to put together, for I did not find any source that would gather all the lyrics in one page/file. Therefore, the lyrics had to be copy-pasted from single song entries at lyricsfreak.com/b/bob+Dylan (the most complete collection, 2018 updated). I wanted to make the copying process easier by using some kind of a data mining programme, however I let go of the idea, because the webpages had too much additional text/ads and it would take some deeper research into datamining to extract the lyrics in the form I needed.

Process

As a very unexperienced person in what concerns programming, I decided to use the tools that I am used to as much as possible. For pre-processing the text, I have used Word, Excel and WordPad and just later on I tried my best with Python.

First, I put both corpora into separate MS Word files and removed the French parts and Introduction of the Rimbaud collection. Then I changed all the letters into lowercase using the built-in Word function for that. I used the Find and Replace tool to remove diacritics (possible in most cases without a problem, but to remove for example a question mark, one has to put a tilde sign in front of it in the command window – "~?")
Because the text I worked with was in the format of poems and lyrics, it was quite easy to transfer it into Excel lines. There I proceeded to remove the duplicate lines of refrains (especially frequent in Dylan's songs) and empty lines.

With both texts in .txt files it was time to try and filter out the stop words (words like "as", "if", "what", "for"), so the results would have a clearer meaning. For removing these I had to install Python (from python.org). I had to install Python twice, because for some reason the script file did not install at the first time. The script contained the pip3 file that was necessary for installing the NLTK package (The Natural Language Toolkit) which can be done in Windows Command Prompt (> pip install nltk). I have used the instructions on the pages stackoverflow.com/questions/49048948/windows-10-python-3-6-install-nltk guru99.com/download-install-nltk.html to guide me through the installation. Getting the Python to work was a struggle, so I tried looking for alternatives and found an application that uses R to filter out the stop words (at https://demos.datasciencedojo.com/demo/stopwords/). I tested it out and found that it works quite well, however is unable to process big data chunks and so I could not use it in this case.

I also explored the possibility to use R coding language, as I stumbled on it while trying to find answers to the various problems I faced when trying to get the Python to work. However in the end Python indeed turned out to be the solution to the task. To define the working file so that Python could find the text files, I followed these instructions: https://stackoverflow.com/questions/1810743/how-to-set-the-current-working-directory

The final code that I used to cleanse the text was:

```
>from nltk.corpus import stopwords
>from nltk.tokenize import word_tokenize

>r = open("Filename.txt","r")
>stop_words = set(stopwords.words('english'))
>word_tokens = word_tokenize(r.read())
>filtered_sentence = [w for w in word_tokens if not w in stop_words]


>for w in word_tokens:
    >if w not in stop_words:
        >filtered_sentence.append(w)

>print(word_tokens)
>print(filtered_sentence)
```

Instructions to set up the stop word analysis, tokenize the text and erase stop words (most of the code) I found also from: https://pythonspot.com/nltk-stop-words/ (https://chrisalbon.com/machine_learning/preprocessing_text/remove_stop_words/ https://www.geeksforgeeks.org/removing-stop-words-nltk-python/)

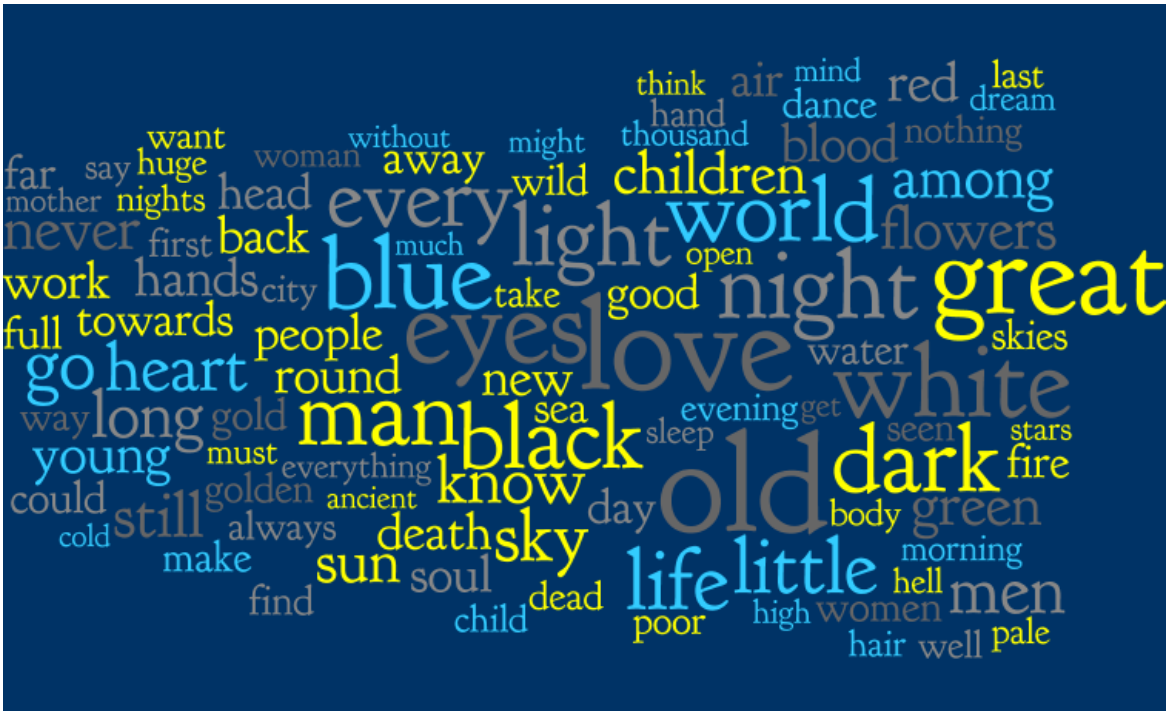
And to open and read the file: https://www.pythonforbeginners.com/files/reading-and-writing-files-in-python.

Through this I got a list of words cleansed from the stop words. These I copied to Word for further processing with the help of Voyant Tools. Loading the data into VT showed remaining stop words (“ah”, “well”) that needed to be removed from both the word clouds. This additional removal was done using the Find and Replace tool in Word.
Voyant Tools links: Dylan https://voyant-tools.org/?corpus=23039b4513814a4608a2c60a8b6d726a , Rimbaud https://voyant-tools.org/?corpus=39ed9ab18c50003a358004b82912793d.

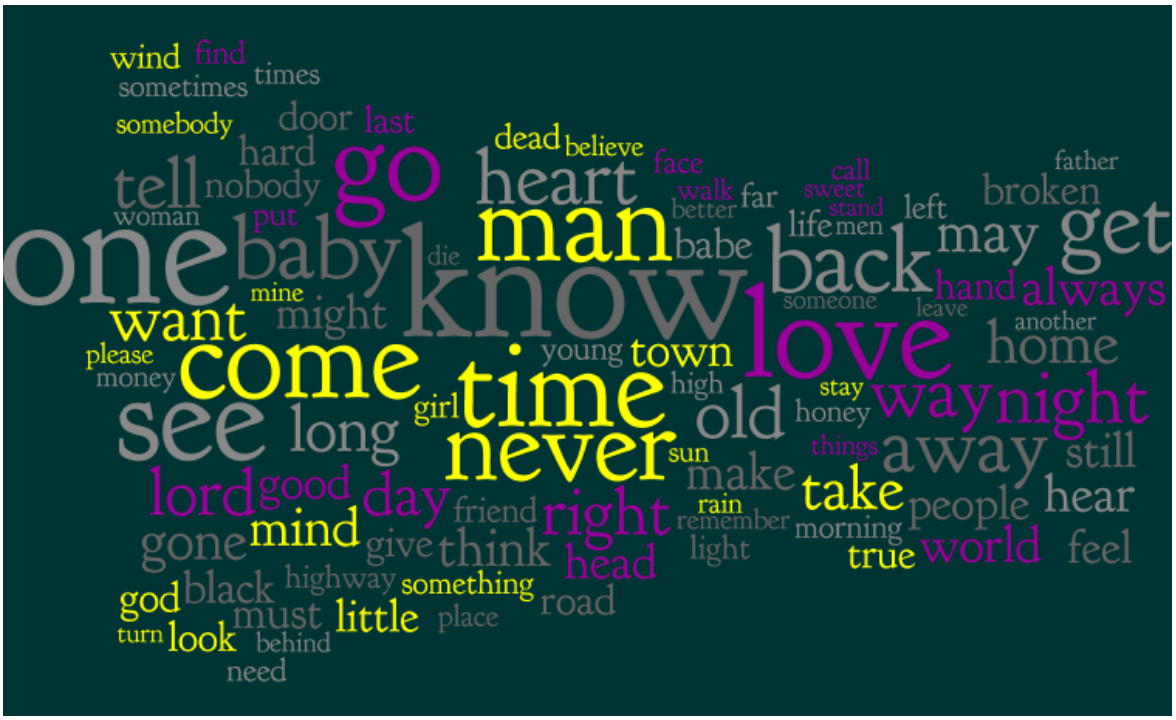
After managing to pre-process the data, I searched for a suitable tool to visualize the results. I have tried various word cloud programs before I settled for Wordle, used in one of the previous final projects. Wordle cannot be used in the browser anymore, so it needed to be

installed (which also was not uneventful, as a couple of errors popped up and missing Java packages had to be found and downloaded). Using Wordle I have created two word clouds of 100 words from the cleansed words.

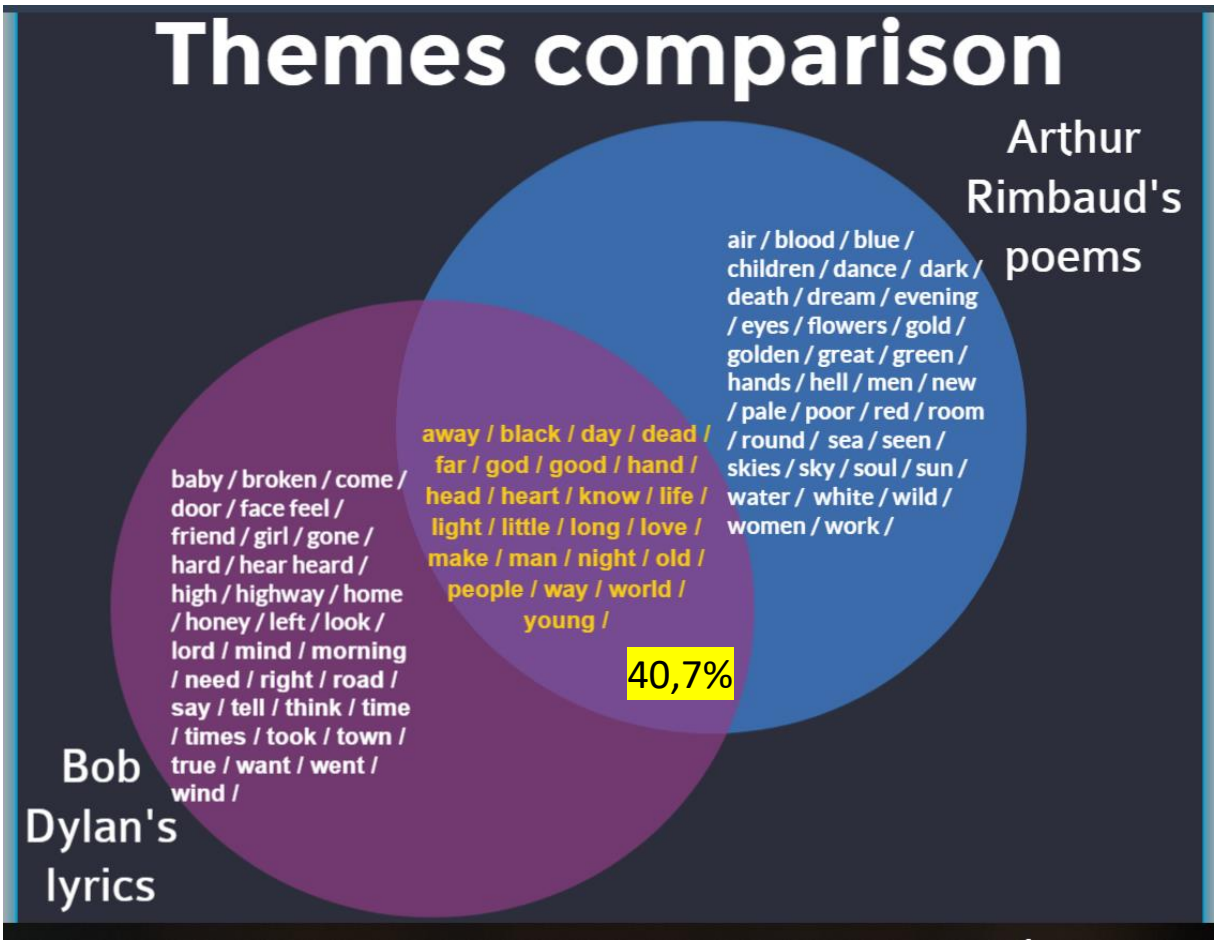
Rimbaud:



Dylan:



To compare the words, I used an online list comparing engine at http://barc.wi.mit.edu/cgi-bin/barc/tools/compare.cgi. This has shown that out of 59 most common words (each), Rimbaud and Dylan share 24 of them (40,7%). I created a diagram using Visme (visme.co).



What does the analysis show?

This analysis shows that over 40% of the main themes in the poems and lyrics in question is shared. However, I do not think it does justice to the connection that Dylan feels to Rimbaud's works. Themes analysis in poems is tricky, for the poet seldom names what he really means, and it is just the same with song lyrics. So, although the similarities can be calculated like this, the overall feeling from experiencing the poems/songs is still more meaningful. Words stripped of context cannot have the power the writer intended them to have.

What I think is interesting is that even if you would not know which side of the diagram is whose, you would figure it out by reading the expressions used. Rimbaud's expressions have

some typically symbolist features – for example using the colour or nature symbolism. The era difference between the two is visible. Dylan's expressions are compared to Rimbaud's more modern ("highway", "baby"), yet in both cases there is a sense of longing for movement, getting away from somewhere "far", "away", "world", "way". They both use basic poetic contrasts of "young" x "old", "day" x "night". They both play with topics of love and death. "Heart", "head" and "hand" are also typical symbols in poetry, synecdoches.
A noticeable feature is Rimbaud's expressions is adding dramatic contrast with "dark", "blood", "hell" – just as expected from an "Accursed poet". Dylan on the other hand sounds more casual and uses familiar expressions like "honey", "baby".
Dylan's inspiration by Rimbaud is undeniable, however it probably is even more on a spirit level than in adopting the very same themes.

Potential problems and bias

As an inspiration for this project I have used the previous projects about Hungarian love songs and Beck's lyrics. Reading through these materials, the task on hand seemed relatively easy, however it proved to be much trickier than expected as I had to wage my own battle with coding. Having a better computer than mine would very surely improve the quality of this project, as most of the problems I encountered and lost time at were due to program installations.

When deciding which words to remove from the lists, past tenses of the verbs that already appear there should have possibly been removed, to make space for more original expressions. Same could be done with plurals of nouns. I decided not to remove verbs and some other expressions bearing a meaning, for I thought they might be interesting in the comparison. Possibly even expressions like "oh" or "ah" could have been kept to pinpoint the individual styles of the writers even better.

What I think is a great problem in analysing poetic texts in this way is the lack of context, as many words can have multiple meanings and those become unclear without the context.

It is also important to point that the English translations of Rimbaud's texts were used in this project and they may of course differ from the French original. However, the English

translation may have been just what Dylan drew his inspiration from (and therefore it would tell more about Dylan than about Rimbaud).

This project has been a great struggle for me, as I have not had any experience with coding or anything similar before. However, I think I have learned some actually useful skills, even concerning programs I have used for a long time like Word and Excel. I'm sure to use and improve my skills in digital humanities in the future.