Isabel Petersen
University of Helsinki
21 December 2024
Computational Literacy

Spatial Relation Extraction from Narrative

1 Topic and research questions

Every story is narrated through a narrative. The narrative is the 'how': how are events represented? That seems fair enough, but within literary theory, the discourse around the definition has not found common ground. Especially when it comes to what the minimal elements that constitute a narrative are. It is generally agreed that a narrative requires temporality within which events take place. However, though it has been recognised that spatiality plays a significant role in narrative, spatial narratology has received substantially less attention than temporal narratology in narrative studies (Ryan 2012). Immanuel Kant argued in 1781 that both time and space are a priori meaning they are not learned experiences but rather pure forms of existence (Ryan 2012). Much of philosophy focuses on the intimate relationship between those two entities and in physics spacetime is of course associated with the Einsteinian theory of relativity. Space and time are dimensions that do not exist without the other and that is also reflected in the stories we tell and understand.

Spatiality has already been approached computationally for a few decades but there are still great issues and complications with detecting and dealing with spatiality in text (see e.g. Manzoor & Kordjamshidi 2018; Kordjamshidi, Rahgooy & Manzoor 2017; Kelleher & Costello 2008). For this project, digital methods were applied for extracting and analysing expressions that indicate spatial relation in narrative. The methods are applied to the text *Alice's Adventures in Wonderland* by Lewis Carroll. The initial research question was:

1. How does spatiality drive the narrative forward?

During the process of the project additional questions arouse:

2. What is place and spatiality? How can spatiality be defined and narrowed down to a concept that can be approached computationally?
3. Where and what kind of Spatial Indicators occur?

2 Defining space and extracting space from the data: problems and bias

Space is not a straightforward concept. In many computational studies of space, space is understood as consisting of three pillars or 'spatial roles':

- Trajectors
  - The entity whose location is described
- Landmark
  - The reference object for describing the location of the trajector
- Spatial indicators
  - Spatial information in text

(see e.g. Kordjamshidi, Rahgooy & Manzoor 2017, 34, https://aclanthology.org/W18-1407). This project aligns itself with the conceptual framework of space consisting of trajectors, landmarks, and spatial indicators. The primary focus is on spatial indicators.

The problem with space and place is that spatial information is complex. Especially the following aspects create issues:

- spatial semantic ambiguities (see Viehhauser-Mery & Barth 2017)
- metaphors (*sink into deep sea* vs. *sink into deep thought*)
- referring back to the same spatial entities/relations with anaphora (e.g. *it, them*) (see Manzoor & Kordjamshidi 2018,)
- hazy boundary when an object become a space/place only after being introduced as a non-spatial object (e.g. a tree is mentioned but later the tree is climbed which makes the tree a place) (see Viehhauser-Mery & Barth 2017)

These issues can create issues both with quantitative and qualitative analysis.

I decided to define my own set of Spatial Indicators that indicate Spatial Relation and ended up with a list consisting of prepositions that can belong to prepositional phrases and/or phrasal verbs as well as the additional expressions ‘back’, ‘here’, and ‘there’ (see figure 1.: Spatial Indicators). I decided to only focus on these linguistic markers as one can argue that Spatial Indicators always connect (to) landmarks and trajectors and therefor fetching only those will already give a good understanding of a given space. When occurring in a text, the Spatial Indicators trigger modification of space by adding or changing elements, so they can be seen as the spatial features that drive the narrative forward.

It has to be mentioned that this way of defining and (out of necessity) narrowing down the definition of space/Spatial Relations cause a bias to the approach. Working with a different set of spatial expressions (e.g. landmarks) can dramatically change the results. Additionally, further bias emerges from manual cleaning of the data which relied on interpretive choices. These choices will be further elaborated on in the next section.

3 Processing and transformation of the data

3.1.1 Text from Project Gutenberg

The novel *Alice's Adventures in Wonderland* by Lewis Carroll, first published in 1865, was used for this project. Spaces and places play a particularly interesting role in this narrative as Alice begins her adventure by following a mysterious rabbit into a rabbit hole and soon after she finds herself in a fantastical world to explore. Intersectional and interdisciplinary studies of space in this exact work have been carried out (Grandy & Tuber 2009; Swamidoss & Tally 2018), but, to my knowledge, only through qualitative analysis.

For this project, I used the first chapter from *Alice's Adventures in Wonderland* with the chapter title "Down the Rabbit-Hole". The text extract was retrieved from Project Gutenberg at https://www.gutenberg.org/ebooks/11.

3.1.2 Yoshikoder

I extracted the first chapter of *Alice's Adventures in Wonderland* from Gutenberg and saved it into a plain .txt file. I imported the text file to the program Yoshikoder (Lowe 2015). Initially, I wanted to use the Regressive Imagery Dictionary (RID). With the help of RID, Yoshikoder can detect words that are associated with spatial meaning, as the dictionary can highlight words belonging to categories like depth, height, ascend and descent (among many other non-spatial categories). However, I ended up creating my own dictionary because I felt the categories were too random and I wanted a neater categorisation. I came up with the idea to only look at spatial indicators that indicate Spatial Relation. In the end, I created the Dictionary "SpatialRelationsDictionary.ykd" with the Categories "Prepositional", "Demonstrative", and "Other" containing the following "patterns" (= the expressions or linguistic markers that the program detects):

category	**pattern: linguistic marker**
Prepositional	across, against, along, at, behind, by, down, from, in, into, near, of, off, on, out, over, through, to, under, underneath, up, with
Demonstrative	here
Other	back

I then used the "make concordances" function in Yoshikoder to get the hits with their surrounding context which is five words to the left and five words to the right of the pattern. Yoshikoder automatically normalises the concordances by removing punctuation marks and replacing apostrophes with blank space. Cases remain unchanged (see Fig. 1).

Now looking back at the process, the data extraction with Yoshikoder could have been replaced with python code but working with Yoshikoder helped me to get started and made it possible to easily explore different patterns. Additionally the user interface makes it easy to see everything in context.

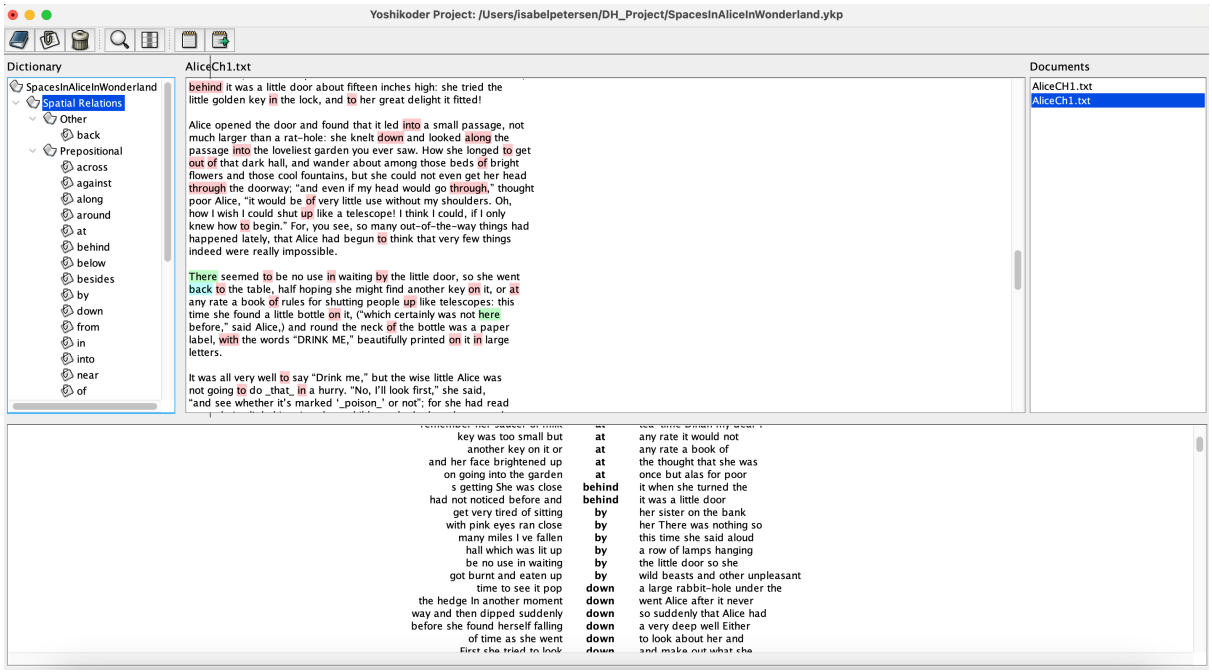

Fig. 1: Screenshot from Yoshikoder's interface.

3.1.3 Export to Excel and cleaning of the data

The concordance list was exported to an Excel sheet. Each hit was reviewed manually to decide whether it truly was expressing Spatial Relation. Interpretative choices about whether to include or exclude were made and therefor the results might be prone to bias. All deletions are documented in the sheet 'cleanedData' of the Excel file 'AliceSpReCh1.xlsx' and a justification for inclusion/exclusion is presented in ambivalent cases (see 'In cases of ambiguity: Explanatory Notes' in the Excel file).

The most important general guidelines I made for this project were:

- space and place that was talked, thought, or dreamt about was kept in addition to space that characters or things were actually situated in at the moment of narration
    - e.g.:

if I shall fall right	through	the earth How funny it

- space connected to the physical body was deleted from the list
    - e.g.

Which way holding her hand	on	the top of her head

- grammatical dummy subject 'there' was deleted from the list
    - e.g.

were down here with me	There	are no mice in the

    - "There" here means "There exist" as opposed to e.g. "the mice are there"

The final list of occurrences of Spatial Relations can be found in the sheet 'finalList' of the Excel file 'AliceSpReCh1.xlsx'. Not all 285 initial hits were kept. After manual cleaning, the final list of Spatial Relations contains the Spatial Indicators: *back, across, along, at, behind, by, down, from, in, into, near, of, off, on, out, through, to, under, underneath, up, here,* and

there (i.e. excluding the Spatial Indicators *against, around, below, besides, outside, over, towards, under, with* either becomes the text did not contain them or because they were used with non-spatial meaning). Altogether that makes only 22 different Spatial Indicators. These various Spatial Indicators occur 104 times altogether in the first chapter meaning only 36,5% of automatically detected occurrences of the list Spatial Indicators actually indicated Spatial Relation (as defined for this project).

3.1.4 Python code

The code is located in the file 'SpaRe.py'. The program locates Spatial Relation in a text and creates a visualisation of the frequency distribution of Spatial Indicators. I used the packages math for math.ceiling (for rounding up numbers), the package pandas (for conversion of Excel to list ls), and matplotlib (for visualisation purposes).

3.1.5 Visualisation

The results were visualised as a bar chart showing the frequency distribution of Spatial Indicators (see Fig. 2). The x-axis presents the whole text extract (10792 characters) segmented into chunks. Here I divided the text into 50 chunks which means each chunk contains 216 characters. The y-axis represents the frequency of Spatial Indicators per segment/chunk.

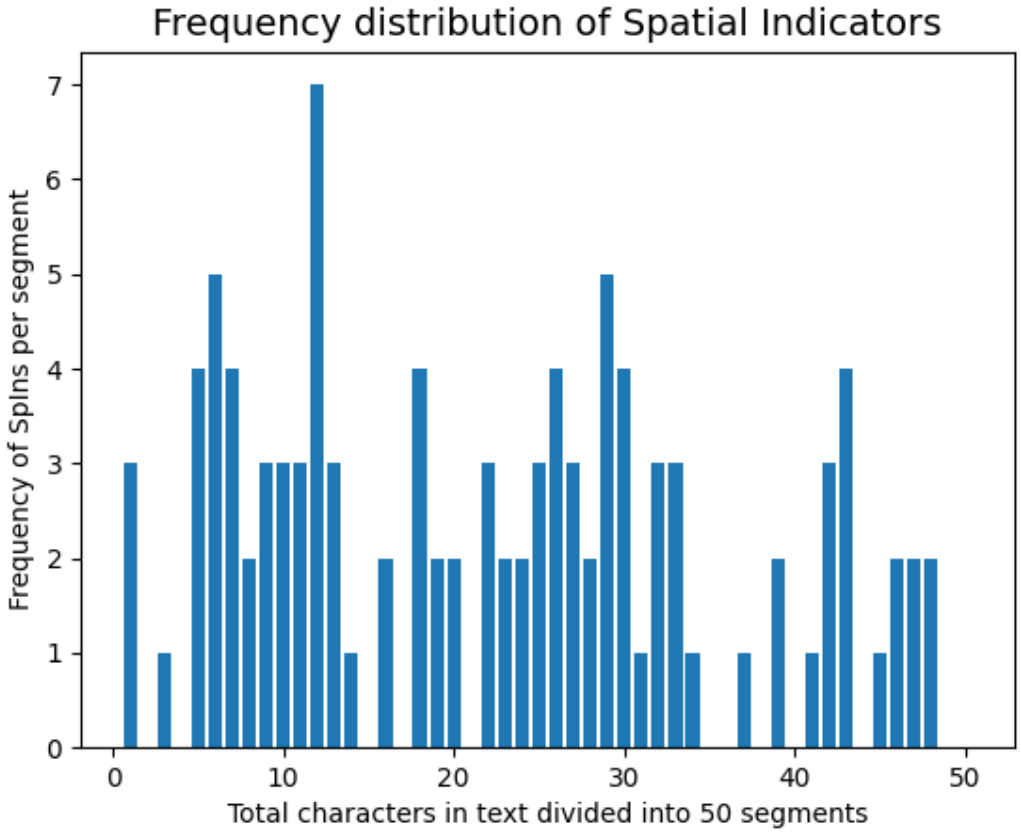

Fig. 2: Frequency distribution of expressions that work as Spatial Indicators in the text.

4 Analysis of results and contextualisation

From the results one can conclude that expressions referring to spatial relation occur almost throughout the text. There are few chunks that do not contain any Spatial Indicators of Spatial Relation but even in these cases there is only about 216 characters which equals about 40 words in between occurrences of Spatial Relation. From the chart one can detect one peak that stands out in chunk twelve (characters 2387 to 2603):

> thing of tumbling down stairs How brave they ll all think me at home Why I wouldn t say anything about it even if I fell off the top of the house Which was very likely true Down down down Would the fall never come to

> [original:] think nothing of tumbling down stairs! How brave they'll all think me at home! Why, I wouldn't say anything about it, even if I fell off the top of the house!" (Which was very likely true.)
> Down, down, down. Would the fall *never* come to

When inspected qualitatively it soon becomes clear why this segment stands out. There is hypothetical structure with spatial information "if I fell off the top of the house" but more importantly, Alice's fall through and down the rabbit hole is described with "Down down down" (in the original " Down, down, down."). This does not come as a big surprise. Already in the manual cleaning of the concordance list I came across this and asked myself whether to include only one of the successive repetitions of this word. However, I decided to keep all of them because the rhetorical repetition can be interpreted as adding emphasis to the degree of deepness and thus being of relevance. This segways into an important aspect of space in fictional as opposed to non-fictional space. In fiction, space does not have to be logical or factually correct. While falling, Alice thinks to herself: "I wonder if I shall fall right *through* the earth! How funny it'll seem to come out among the people that walk with their heads downward!". Nevertheless, a reader will have a mental visualisation of what the mentioned scene would look like spatially.

One of the second highest frequencies is the 29th chunk (characters 6076 to 6292):

> golden key in the lock and to her great delight it fitted Alice opened the door and found that it led into a small passage not much larger than a rat-hole she knelt down and looked along the passage into the lovelies

> [original:] golden key in the lock, and to her great delight it fitted! Alice opened the door and found that it led into a small passage, not much larger than a rat-hole: she knelt down and looked along the passage into the loveliest

Alice has just entered Wonderland and thus, the places and spaces she finds herself in become central. The narration is very much guided by where Alice is, where she comes from, and where she is headed. One will find that the narration focuses comparatively less on time. New elements are introduced through spatial relations like “led into a small passage” which guide the narrative forward. A reader understands the narration by creating a mental conceptualisation of the described space. The living handbook of narratology describes: “Mental maps, in other words, are both dynamically constructed in the course of reading and consulted by the reader to orient himself in the narrative world.” (Ryan 2014).

My project aimed at investigating space in narrative. The most important outcomes are that spatial relations are continuingly present in the text extract of *Alice’s Adventures in Wonderland.* These spatial relations work to both situate characters and objects into the fictional world and build relationships between entities. Spatial relations can also introduce new objects, characters, and places. The digital methods used yielded interesting results, though the automatic detections of spatial relation required a great amount of qualitative analysis to clean erroneous data. In that sense, the tools would need to be refined in order to analyse larger text extracts or corpuses and subsequently be able to make more generalisable conclusions.

References

Manzoor, U., & Kordjamshidi, P. (2018). Anaphora resolution for improving spatial relation extraction from text. In Proceedings of the First International Workshop on Spatial Language Understanding (pp. 53–62). New Orleans: Association for Computational Linguistics.

Ryan, M.-L. (2014). Space. In P. Hühn et al. (Eds.), The living handbook of narratology. Hamburg: Hamburg University. Retrieved from http://www.lhn.uni-hamburg.de/article/space.

Kordjamshidi, P., Rahgooy, T., & Manzoor, U. (2017). Spatial language understanding with multimodal graphs using declarative learning-based programming. In Proceedings of the 2nd Workshop on Structured Prediction for Natural Language Processing (pp. 33–43). Copenhagen: Association for Computational Linguistics.

Kelleher, J., & Costello, F. (2009). Applying computational models of spatial prepositions to visually situated dialog. Computational Linguistics, 35(2), 271–306. https://doi.org/10.1162/coli.06-78-prep14

Grandy, M. A., & Tuber, S. (2009). Entry into imaginary space: Metaphors of transition and variations in the affective quality of potential space in children’s literature. Psychoanalytic Psychology, 26(3), 274–289. https://doi.org/10.1037/a0016446

Swamidoss, H., & Tally, R. T. (2018). Teaching non-places in British children's fantasy literature. In Teaching Space, Place, and Literature (1st ed., pp. 208–218). Routledge. https://doi.org/10.4324/9781315171142-22

Viehhauser-Mery, G., & Barth, F. (2017). Towards a digital narratology of space. Digital Humanities Conference.

Lowe, W. (2015). Yoshikoder: Cross-platform multilingual content analysis (Version 0.6.5) [Software]. Retrieved from https://yoshikoder.org