

---

# Towards a Digital Narratology of Space

Gabriel Viehhauser-Mery  
viehhauser@ilw.uni-stuttgart.de  
Universität Stuttgart, Germany

Florian Barth  
florianbarth@ilw.uni-stuttgart.de  
Universität Stuttgart, Germany

---

## Introduction

Besides time, characters and plot, space is one of the main components in storytelling. But despite its importance as a category for the setting of narrative action and unlike the other mentioned categories, the conceptualisation of space has long been neglected in narratological research. This holds true even after the so-called *spatial turn* (Soja 1990) in cultural history, that lead to a renewed interest and to fruitful insights into space as a metaphorical concept. However, a systematic description of the means by which space is created in narratives is still in its beginnings (e.g. Dennerlein 2009, Piatti et. al. 2009).

This is at least partly due to the fact that space poses substantial problems for modeling. The creation of space in narratives is often dynamic and based on implicit information: Rather than constructing a given, mathematical space beforehand, stories tend to evolve their setting in relation to its characters that constitute space through their actions. Spatial information in stories therefore highly depends on the characters that act, move or perceive within it. Especially in fiction, this also means that spatial information is often fuzzy and imprecise (Piatti et. al. 2009), since narrators quite frequently are more interested in telling a story than designing a detailed, coherent setting for it. Whereas these problems are hard to handle in traditional literary studies, they present serious yet interesting challenges for a digital formalization.

In our paper, we will illustrate the complex tasks that have to be tackled by a digital narratology of space based on an exemplary annotation workflow, that we will outline for the description of spatial elements in Jules Verne's *Around the world in Eighty Days*.

## Challenges for a digital narratology of space

We describe the following major problems that can be grouped into a chain of work-tasks:

1. Basic information on the setting of a narrative can be retrieved by extracting the place names from a text (NER). However, an automatic extraction is flawed by well-known problems of disambiguation (cf. for example, Barbaresi / Biber 2016). Since space in narration is highly dependent on characters that are placed in it, these entities have to be detected as well.
2. Place names are not the only kind of spatial information that can be found in texts. Besides others, space is also constituted with the help of nouns that do not necessarily have an inherent spatial component (for example, a *car* in a text can be the subject of a description, but it turns into a space marker if someone enters it).
3. Spatial entities (names and nouns) can be referred to by co-reference.
4. Spatial entities in a text are not always the setting of narrative actions. Place names or nouns can also only be mentioned, dreamed of, remembered, reflected on etc. This different functionality has to be taken into account when it comes to the automatic generation of literary maps (eg. Moretti 1998, Piatti 2008). To capture this opposition, Dennerlein (2009) separates *event regions* from *mentioned spatial objects* in her conception. *Event regions* are defined as spatial zones, where events take place. In contrast, *mentioned spatial objects* contain all spaces that are not event-related. Piatti et. al. (2009) develop a similar model: Their concept of *setting* closely corresponds with *event regions*, whereas *projected space* and *marker* give a finer differentiation of the notion of *mentioned spatial objects* (cf. Figure 1).

Dennerlein	Piatti
<b>Event regions</b> <ul style="list-style-type: none"><li>- Spatial zone, where an event takes place</li><li>- The event determines the extension of the place (wide focus)</li></ul>	<b>Setting</b> <ul style="list-style-type: none"><li>- Characters need to be present</li><li>- Each single action of a character at a current setting constitute this kind of place (narrow focus)</li></ul>
<b>Mentioned spatial objects</b> <ul style="list-style-type: none"><li>- Contains all spaces that are not part of an event</li><li>- This includes commenting, arguing, reflections or descriptions</li></ul>	<b>Projected space</b> <ul style="list-style-type: none"><li>- Characters are not present in this location, but they dream of, remember, or long for it</li></ul> <b>Marker</b> <ul style="list-style-type: none"><li>- Describes a location that is only mentioned</li><li>- It has no significance for the story or the character</li><li>- Markers indicate the geographical range and horizon of the fictional space</li></ul>

Figure 1: Comparison of spatial concepts from Katrin Dennerlein Barbara Piatti

## Annotation Workflow

token ID	tokens	POStags	lemmas	namedEntities	wordlists	Dennerlein	co-reference	ISO-Space
L_1	in	IN	in					
L_2	which	WDT	which					
L_3	Phileas	NNP	phileas	B-PER				Spatial_Entity(D=na1, type=PERSON, form=NAME)
L_4	Fogg	NNP	Fogg	B-PER				Spatial_Entity(D=na1, type=PERSON, form=NAME)
L_5	descends	VBZ	descend					Motion(D=na1, motion_type=COMPOUND, motion_class=FOLLOW, motion_sense=LITERAL)
L_6	the	DT	the					
L_7	whole	JJ	whole					Measure(D=me1, value='whole length', unit=d)
L_8	length	NN	length					
L_9	of	IN	of					
L_10	the	DT	the	B-ER		B-Antecedent		
L_11	beautiful	JJ	beautiful			I-Antecedent		
L_12	valley	NN	valley	B-LSC		I-Antecedent		Path (D=na1, type=VALLEY, form=NAME)
L_13	of	IN	of			I-Antecedent		
L_14	the	DT	the			I-Antecedent		
L_15	Ganges	NNP	ganges	B-LOC		I-Antecedent		Path (D=na2, type=BODYOFWATER, form=NAME)
L_16	without	IN	without					
L_17	ever	RB	ever					
L_18	thinking	VBG	think					
L_19	of	IN	of					
L_20	seeing	VBG	see					
L_21	it	PRP	it			B-Anaphor		Spatial_Entity(D=na2, form=NAME) Metalek (id=meta1, objectID1=na1, objectID=na2, refType=REFERENCE)
L_22	.	.	.					Movelink(D=na1, trigger=1, mover=na1, goal=na1)

The complexity of spatial information demands for a multi-faceted approach. Figure 2 shows a spreadsheet with the beginning of chapter 14 of Verne's *Around the World in Eighty Days* that has automatically, semi-automatically and manually been enriched with multiple layers of annotation.

First, the text was tokenized, lemmatized and POS-tagged (columns 1-4).

Second, Named Entity Recognition (NER) was applied to the text (columns 5, both steps were performed with Weblicht [2012]). The NER also identifies the names of the characters. The results of the NER have to be corrected manually. To improve the automation of this step the exploitation of other toponymical bases like [Geonames](#) and [OpenStreetMap](#) will be discussed.

Thirdly, to generate column 6, we used theme-specific wordlists that we built on the base of existing lexicological ontologies (*GermaNet* for German Texts [Hamp/Feldweg 1997, Henrich/Hinrichs 2010]. English Wordlists as shown were provisionally generated manually, but can be built in a similar way). These wordlists were used to automatically tag the text. In Fig. 1, 'valley' has been annotated as *LSC*, which means that it belongs to the word field *landscape*. So far, we created wordlists for *landscape* and *architecture* (in German), which cover a high amount of place nouns. The annotation can be manually supplemented and new wordlists can be created.

Fourthly, in column 7, 'the beautiful valley of the Ganges' was (manually) annotated as *event region* according to the model of Dennerlein (2009). An automatic differentiation between *event regions* and *mentioned spatial objects* will be a challenging task. However, we consider a rule-based extraction of dependency paths to approach the problem

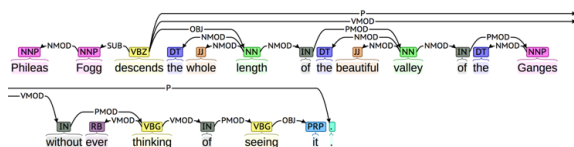


Figure 3 shows a parse tree of the sentence from Verne's text. The pattern [Character – SUBJ – Verb of Motion – OBJ – place noun] is likely to indicate an *event region*. By gaining several similar patterns with high precision regarding the identification of *event regions*, we assume to collect features for a future implementation of machine learning methods.

Fifthly, in column 8, coreferences were annotated manually. We will consider different kinds of co-references: Spatial entities can be referred to by nouns (e.g. 'Paris / 'city') or pronouns. Also, certain deictics ('here', 'there') might refer to spatial antecedents. However, a reliable automatic coreference resolution, which would be highly desirable for many kinds of narratological analysis, is out of the scope of this paper.

Finally, we included a column (9) for annotations that are based on the modified tag-set of the ISO-Space-standard (Pustejovsky et. al. 2011a und b), as they are presented in the SpaceEval Annotation Guidelines (2014).

## Visualizations and Outlook

With the help of this semi-automatic and multi-layered method, we hope that we can make use of the strength of the different approaches and combine their advantages (it would be highly beneficial, for example, to combine Dennerlein's category with an ISO-space-annotated text to enhance their rule-based detection).

The potential of their combination shall be demonstrated by two examples of visualizations that draw on named entities and wordlists.

### A network of spatial markers

As Piatti et. al. (2009) pointed out, the imprecision and semantic potential of spatial information in literary texts sometimes demand visualizations other than geographical maps.

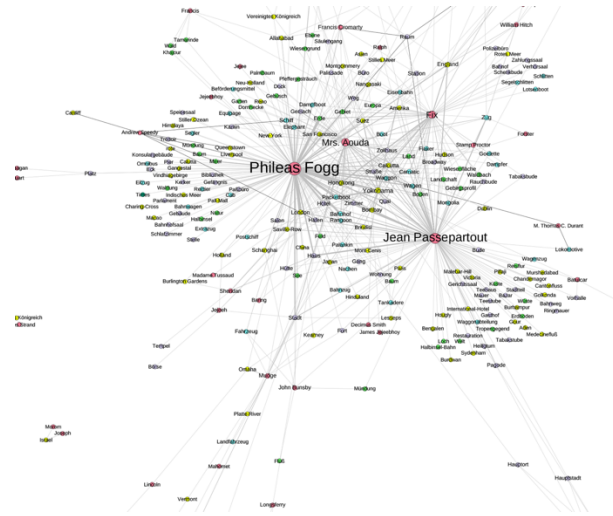


Figure 4 shows a co-occurrence network of characters and place markers in *Around the world in Eighty Days*: Characters appear in red, place names in yellow. Place nouns are divided into the sub-categories *landscape* (green), *architecture* (grey) and *transport* (blue). In a straightforward approach, we established edges whenever a character and a spatial marker appear in the same sentence. The nodes have been sized according to their degree (the number of their connections), which can be related to Juri Lotman's (Lotman 1977) concept of mobile vs. immobile characters: Characters which are connected to many places like Phileas Fogg and Passepartout are more likely to be main characters than characters with a lesser degree.

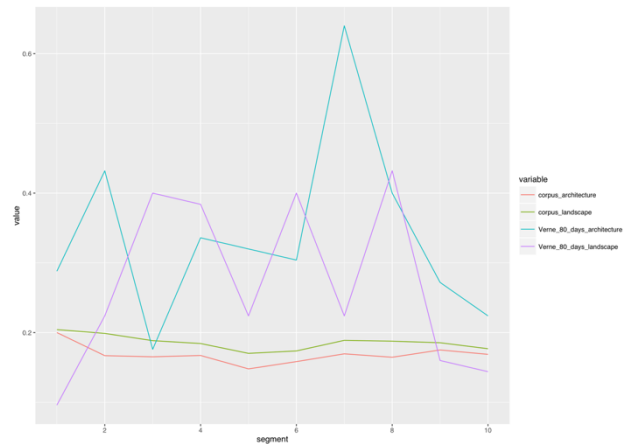
(The visualization was established with Gephi [Bastian et al. (2009)]).

### Word-list-based Frequency Analyses

Figure 5 shows the distribution of *landscape* and *architecture* terms over the whole text of *Around the world in Eighty days* compared to a corpus of 451 German novels taken from the TextGrid-Repository (TextGrid Konsortium 2006-2014, licensed CC-BY-4.0), which cover a time range from 1700 to 1920.

Every text was chunked into 10 'segments' (x-axis), for which we calculated the relative percentage of the vocabulary from the corresponding word field ('value', y-axis). The graph shows a noticeable peak in the use of architectural vocabulary in the last third of the text, which can serve as a starting-point for a *close reading* of the text. However, to take full advantages of *distant reading* techniques for spatial analysis, more refined methods and annotated corpora are necessary. We hope that these methods can be developed by considering the challenges outlined in our basic model in this

paper.



### Bibliography

- Barbaresi, A., and Biber, H.** (2016): „Extraction and Visualization of Toponyms in Diachronic Text Corpora.“, in: Digital Humanities 2016, Jul 2016, Cracovie, Poland, Digital Humanities 2016 Conference Abstracts, 732-734 <http://dh2016.adho.org/>
- Bastian, M., Heymann, S., and Jacomy, M.** (2009): *Gephi: an open source software for exploring and manipulating networks*. International AAAI Conference on Weblogs and Social Media.
- Dennerlein, K.**(2009): *Narratologie des Raumes*. Berlin: de Gruyter.
- Hamp, B., and Feldweg, H.**(1997): „GermaNet - a Lexical-Semantic Net for German.“, in: Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications. Madrid
- Henrich, V., and Erhard, H.** (2010): "GernEdiT - The GermaNet Editing Tool", in: Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010). Valletta, Malta, May 2010, 2228-2235.
- Lotman, J.** (1977): *The Structure of the Artistic Text. Translated from the Russian by Ronald Vroon*. Ann Arbor: University of Michigan, Department of Slavic Languages and Literatures.
- Moretti, F.** (1998): *Atlas of the European novel. 1800-1900*. London / New York: Verso.
- Piatti, B.** (2008): *Die Geographie der Literatur. Schauplätze, Handlungsräume, Raumphantasien*. Göttingen: Wallstein.

**Piatti, B., Bär, H. R., Reuschel, A.-K., Hurni, L., Cartwright, W.** (2009): „Mapping Literature: Towards a Geography of Fiction.“, in: Cartwright, William / Gartner, Georg / Lehn, Antje (Eds.): *Cartography and Art*. Berlin / Heidelberg, Springer 2009, 179-194.

**Pustejovsky, J., Moszkowicz, J. L., Verhagen, M.** (2011a): ISO-Space: The annotation of spatial information in language, in: *Proceedings of the Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, 1–9.

**Pustejovsky, J., Moszkowicz, J. L., Verhagen, M.** (2011b): Using ISO-Space for Annotating Spatial Information. In: *Proceedings of the International Conference on Spatial Information Theory*

**Soja, E.** (1990): *Postmodern Geographies. The Reassertion of Space in Critical Social Theory*. London / New York: Verso.

**SpaceEval Annotation Guidelines** (2014)  
<http://jamespusto.com/wp-content/uploads/2014/07/SpaceEval-guidelines.pdf>

**TextGrid Konsortium** (2006–2014). *TextGrid: Virtuelle Forschungsumgebung für die Geisteswissenschaften*. Göttingen: TextGrid Konsortium. [textgrid.de](http://textgrid.de).

**WebLicht** (2012): CLARIN-D/SfS-Uni. Tübingen 2012. WebLicht: Web-Based Linguistic Chaining Tool. Online. <https://weblicht.sfs.uni-tuebingen.de/>