# A Visual Representation of Wittgenstein's *Tractatus Logico-Philosophicus*

**Anca Bucur**
Center of Excellence in Image Study,
Faculty of Letters,
Solomon Marcus Center
for Computational Linguistics,
University of Bucharest
anca.m.bucur@gmail.com

**Sergiu Nisioi**
Faculty of Mathematics and
Computer Science,
Solomon Marcus Center
for Computational Linguistics,
University of Bucharest
sergiu.nisioi@gmail.com

## Abstract

In this paper we present a data visualization method together with its potential usefulness in digital humanities and philosophy of language. We compile a multilingual parallel corpus from different versions of Wittgenstein's *Tractatus Logico-Philosophicus*, including the original in German and translations into English, Spanish, French, and Russian. Using this corpus, we compute a similarity measure between propositions and render a visual network of relations for different languages.

## 1 Introduction

Data visualization techniques can be essential tools for researchers and scholars in the humanities. In our work, we propose one such method that renders concepts and phrases as a network of semantic relations. In particular, we focus on a corpus built from different translations of the *Logisch-Philosophische Abhandlung* (Wittgenstein, 1921) from German into English, French, Italian, Russian, and Spanish.

Wittgenstein in his later works states that *meaning is use* (Wittgenstein, 1953): *43. For a large class of cases –though not for all– in which we employ the word "meaning" it can be defined thus: the meaning of a word is its use in the language game. And the meaning of a name is sometimes explained by pointing to its bearer.*

This idea anticipated and influenced later research in semantics, including the *distributional hypothesis* (Harris, 1954; Firth, 1957) and more recently, work in computational linguistics (Lenci, 2008). Distributional semantics works on this very principle, by making use of data to build semantic structures from the contexts of the words. Word embeddings (Mikolov et al., 2013) are one such example of semantic representation in a vector space constructed based on the context in which words occur. In our case, we extract a dictionary of concepts by parsing the English sentences and we infer the semantic relations between the concepts based on the contexts in which the words appear, thus we construct a semantic network by drawing edges between concepts.

Furthermore, we generalize on this idea to create a visual network of relations between the phrases in which the concepts occur. We have used the multilingual parallel corpora available and created networks both for the original and the translated versions. We believe this can be helpful to investigate not only the translation from German into other languages, but also how translations into English influence translations into Russian, French or Spanish. For example, certain idioms and syntactic structures are clearly missing in the original German text, but are visible in both the English and Spanish versions.

## 2 Dataset

The general structure of the text has a tree-like shape, the root is divided into 7 propositions, and each proposition has its own subdivisions and so on and so forth, in total numbering 526 propositions. A *proposition* is the structuring unit from the text and not necessarily propositions in a strict linguistic sense. Our corpus contains the original German version of the text (Wittgenstein, 1921) together with translations into 5 different languages: English, Italian, French, Russian, and Spanish. For English, we





*Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*,
pages 71–75, Osaka, Japan, December 11-17 2016.

have two translations variants, one by Ogden and Ramsey (1922) revised by Wittgenstein himself and another one by Pears and McGuinness (1961).

Since the text has a fixed form structure, it is straight forward to align each translation at the proposition level. In addition, we also employ a word-alignment method to create a multilingual parallel word-aligned corpus and to be able to inspect how certain concepts are translated into different languages. The exact size of each version in the corpus[1] is detailed in Table 1. Our corpus contains a relatively small number (526) of aligned examples and alignment methods often fail to find the correct pairs between words. To create the word-alignment pairs, we have experimented with different alignment strategies including GIZA++ (Och and Ney, 2000), fast align (Dyer et al., 2013) and efmaral (Östling and Tiedemann, 2016), while the later proved to output the best results in terms of our manual evaluation.

| Language | Translator | No. of tokens | No. of types |
|---|---|---|---|
| German | —— | 18,991 | 4,364 |
| English | Ogden and Ramsey | 20,766 | 3,625 |
| English | Pears & McGuinness | 21,392 | 3,825 |
| French | G.G. Granger | 22,689 | 4,178 |
| Italian | G.C.M. Colombo | 18,943 | 4,327 |
| Russian | M.S. Kozlova | 10,682 | 4,090 |
| Spanish | E.T. Galvan | 13,800 | 3,191 |

Table 1: The size of each corpus in the dataset

The two translations into English share a lot in common, however they are not equivalent, for example, the German concept *Sachverhaltes* is translated by Ogden and Ramsey (1922) as *atomic facts* and in Pears and McGuinness (1961)'s version the same concept is translated as *states of affairs*. As for the other languages, the Spanish and Russian translations resemble more the former English version, *Sachverhaltes* being translated as *hechos atomicos* and атомарного факта *(atomarnogo fakta)*, respectively. In French and Italian, the concept is translated as *états des choses* and *stati di cosi* following the Pears and McGuinness (1961) English translation.

## 3 Wittgenstein's Network

### 3.1 Tractatus Network

The *Tractatus Network*[2] is obtained from different versions of the text by computing a pair-wise similarity measure between propositions. Each proposition is tokenized and each token is stemmed or lemmatized. The lemmatizer is available only for English by querying WordNet (Fellbaum, 1998), for the remaining languages different Snowball stemmers are available in NLTK (Bird et al., 2009). Stop words from each proposition are removed before computing the following similarity score:

$$Similarity(p_1, p_2) = \frac{|p_1 \cap p_2|}{max(|p_1|, |p_2|)} \quad (1)$$

The similarity score computes the number of common tokens between two propositions normalized by the length of the longest proposition, to avoid bias for inputs of different lengths. Two propositions are connected by an edge if their similarity exceeds the 0.3f threshold. To render the network, we use a browser-based drawing library[3], the lengths of the edges are determined by the similarity value and the nodes representing propositions are colored based on the parent proposition (labeled from 1 to 7). Furthermore, we added a character n-grams search[4] capability for the network that highlights the node with the highest similarity to the search string.

[1]The dataset is available upon request from the authors.

[2]The *Tractatus Network* is accessible at `https://tractatus.gitlab.io`

[3]`http://visjs.org/`

[4]`http://fuse.js/`

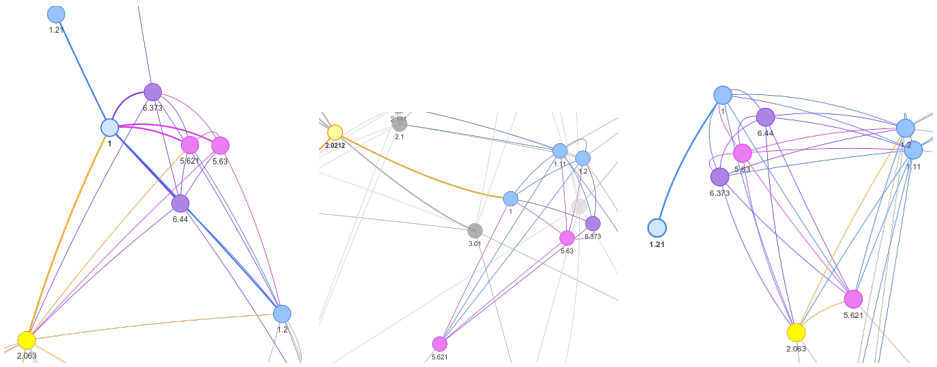

Figure 1: Two excerpts from the *Tractatus Network*. From left to right we have the German original, the translations into English by Pears and McGuinness (1961) in the center, and the Ogden and Ramsey (1922) translation on the right. Propositions from different groups may resemble each other more than the propositions within the same group.

By analyzing the resulted networks, we can observe that the seven main propositions in the text including the sub-divisions are not necessarily hierarchical, at leas not based on the topics addressed, rather the *Tractatus* has a rhizomatic structure in which the propositions are entangled and repeatedly make use of similar concepts. The excerpts rendered in Figure 1 and Figure 2 bring further evidence to this observation, as an example the proposition *die gesamte Wirklichkeit ist die Welt* meaning *the total reality is the world* appears in almost every version close to the propositions in group one in which *die Welt / the world* plays a central role. In Figure 1, the Pears and McGuinness (1961) English translation has a smaller number of relations between propositions, compared to the German counterpart on the left, and it also has an additional proposition from group two: *2.0212 In that case we could not sketch any picture of the world (true or false)*. However, in terms of topology, the Ogden and Ramsey (1922) translation resembles almost identically the German version.

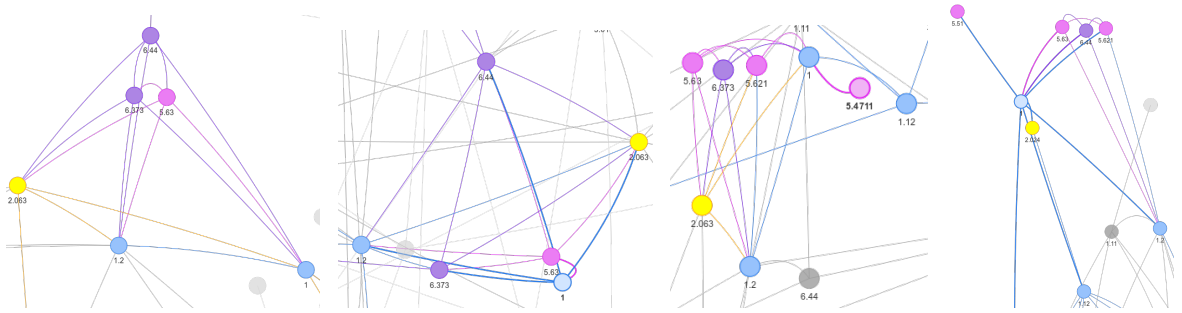

Figure 2: From left to right: Italian, Spanish, French, and Russian excerpts showing the neighbors of proposition 1. Italian and Spanish parts have identical nodes. The French and Russian topologies do not resemble the original or any other network.

On the one hand, looking at the remaining translations, we can observe the Italian and Spanish excerpts share the same nodes and comparable topologies with the original German version. On the other hand, by looking at the word aligned pairs and the translation of *Sachverhaltes* in particular, we may be able trace two separate influences for Spanish and Italian that stem from the different English versions of the *Tractatus*. Last but not least, the French and Russian parts reveal some particularities that cannot be traced to any other topology from the corpus.

It is well known that Wittgenstein did not write the propositions in the order they appear in the text and our results further evidence this fact by revealing specific clusters of similarity between propositions that do not belong to the same group. However, some groups of propositions do appear to be more compact than others, e.g. groups 4 and 2 usually have a more compact structure regardless of the language.

### 3.2 Concept Network

The *Concept Network*[5] is created from the main concepts/keywords extracted from each proposition in the corpus. For this part, we use only the Ogden and Ramsey (1922) translation into English, each proposition is split into sentences and the parse trees are extracted using the approach of Honnibal and Johnson (2015).

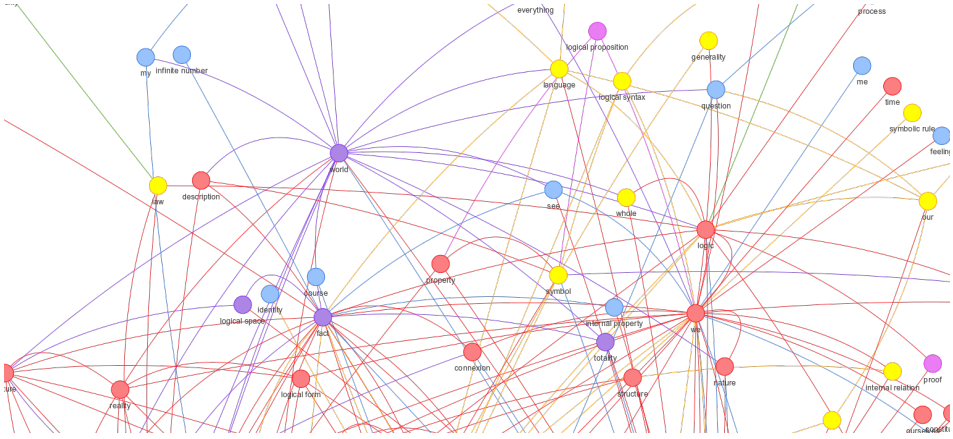


Figure 3: Excerpt from the concept network. The colors indicate the first group proposition in which the concepts appear (from 1 to 7).

The concept list consists of the noun-phrases extracted from the parse trees together with a few personal pronouns that appear in the corpus. We manually pruned the occurrences having low frequencies and the ones that have been wrongly annotated by the parser. The edges between the nodes (concepts) are created based on the number of times a concept appears in at least two propositions in the same context window, where the window varies depending on how many tokens a concept has. Multi word units are allowed to appear in windows of up to ten words, while single token concepts are limited to a maximum window of three words.

An excerpt from the network is rendered in (Figure 3). We noticed that concepts with a high number of edges usually occupy a central position in Wittgenstein's philosophy. Words such as: *elementary proposition*, *proposition*, *world*, *fact*, *form*, *we*, *logic*, *picture*, reveal relations that span across multiple propositions in the text.

## 4 Conclusions

We provide two resources which we believe to be important for scholars and researchers in digital humanities. The first resource is a compiled, word-aligned corpus extracted from the original and translated versions of Wittgenstein's *Tractatus Logico-Philosophicus*. This corpus may be used to study the original text or to extract meaningful comparisons from translations into other languages. The second resource is a web application that renders semantic networks of concepts and propositions from the *Tractatus*. These could be useful to visualize the semantic similarities between concepts and to examine the relations between different propositions, to clarify certain concepts and to search and explore the actual text, either in German or in translation. To summarize, therefore, we hope to provide another method of reading Wittgenstein's work.

## Acknowledgements

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS/CCCDI – UEFISCDI, project number PN-III-P2-2.1-53BG/2016, within PNCDI III

[5]The *Concept Network* is accessible at `https://wittgenstein-network.gitlab.io`

## References


Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media, Inc.

Dyer, C., Chahuneau, V., and Smith, N. A. (2013). A simple, fast, and effective reparameterization of ibm model 2. In *Proceedings of NAACL-HLT*, pages 644–648.

Fellbaum, C. (1998). *WordNet*. Wiley Online Library.

Firth, J. R. (1957). *A synopsis of linguistic theory, 1930–1955*. Blackwell.

Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3):146–162.

Honnibal, M. and Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1373–1378, Lisbon, Portugal. Association for Computational Linguistics.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian journal of linguistics*, 20(1):1–31.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Och, F. J. and Ney, H. (2000). Giza++: Training of statistical translation models.

Ogden, C. and Ramsey, F. (1922). *Wittgenstein, L. - Tractatus Logico-Philosophicus*. Kegan Paul Ltd.

Östling, R. and Tiedemann, J. (2016). Efficient word alignment with Markov Chain Monte Carlo. *Prague Bulletin of Mathematical Linguistics*, 106. To appear.

Pears, D. and McGuinness, B. (1961). *Wittgenstein, L. - Tractatus Logico-Philosophicus*. Classics Series. Routledge.

Wittgenstein, L. (1921). *Logisch-Philosophische Abhandlung*. Annalen der Naturphilosophie, 14.

Wittgenstein, L. (1953). *Philosophical Investigations*. Basil Blackwell, Oxford.