Network Analysis Between Distant Reading and Close Reading

Frederik Elwert

Ruhr University Bochum, Germany
frederik.elwert@rub.de

Abstract

The advent of "distant reading" methods has created the opportunity to look at texts in a new way. But with the shift from close to distant reading, there is also a danger of loosing sight of fine-grained text structure. Like any method, distant reading methodology is not theoretically neutral, but carries a bundle of presuppositions. In this paper, a new method of text network analysis is proposed as a bridge between close and distant reading. Network analysis as a methodological basis offers two advantages: Firstly, it links methodology to theoretical perspectives that highlight relational aspects of linguistic, historical and social phenomena. Secondly, it allows humanities scholars to participate in the interdisciplinary field of network research and to benefit from available methods and tools. The proposed methodology of text-based network generation links close and distant reading in two ways. By building on syntactical structures, the networks resemble closely the linguistic structure of text. And by linking network data and text passages, the method allows to go back and forth between a distant view of the text and its close reading. The method is discussed using Novalis' essay from 1799 "Die Christenheit oder Europa" as an example.

1 Introduction

In the humanities, the availability of large digitised corpora allows for new approaches to the analysis of texts. Scholars have classically read certain selected works with an emphasis on detail, an approach often referred to as "close reading". By applying digital methods, new ways of reading become possible. Those approaches that deal with the analysis of large-scale, quantitative characteristics of corpora have been referred to as "distant reading", a term prominently introduced by Franco Moretti (2013). One of the analytical tools deployed by Moretti and others is network analysis. In its capability to represent, visualise and analyse large-scale structures of texts, network analysis can be seen as a specific way of "distant reading". However, the analytical value of these methods depend largely on the process through which the networkes are derived from the original source. In this paper, I want to describe a new method for the creation of text-based networks that re-introduces elements of close reading. By taking linguistic structures into account, I will argue, network analysis can serve as a bridge between close and distant reading.

The described method has been developed in the research project "Semantic and social network analysis as a means to study religious contact" (short SeNeReKo), a joint project of Ruhr-University Bochum and the Trier Center for Digital Humanities. The project brings together scholars of religion and computer scientists in an effort to develop new tools for the study of aspects of interreligious contact. The project uses two exemplary corpora, the ancient Indian Pali canon and the Thesaurus Linguae Aegyptiae (Berlin-Brandenburg Academy of Sciences and Humanities 2014), a collection of ancient Egyptian texts.

For the purpose of this paper, I will not use these corpora, as the main argument is about the applied methodology, not the content of the historical religious sources. For the sake of demonstration, I will use an essay by Novalis titled "Die Christenheit oder Europa". With this choice, I hope to connect my background in the study of religions with the research area of the conference host, the "Centre Virtuel de la Connaissance sur l'Europe". Novalis' text, written in 1799, is a significant example for the romantic view on modern Europe and the discourse on religion at his time (cf. e.g. Braungart 2011).[1]

1.1 Why network analysis?

Network analysis is appealing to the humanities for two reasons: Firstly, the emphasis on relations between entities matches important theoretical developments. In history, approaches like "histoire croissée" (Werner, Zimmermann 2003),"entangled history" (Conrad, Randeria 2002; also Osterhammel 2001) stress translocal relations as driving forces of historical processes. In sociology, a "relational sociology" is seen as a new paradigm in the study of societies (Emirbayer 1997; Mützel, Fuhse 2010). And in religious studies, contact between religious traditions is now seen as a central aspect for the emergence of religious identities (Krech 2012). These theoretical frameworks can help scholars of religion to question the idea of "world religions" as stable, monolithic entities without neglecting their social impact. By including translocal and transcultural exchange into the study of historical and temporary religions, they can be conceptualized as dense areas in a global religious field (Krech 2012). These theoretical developments do not necessarily come along with elaborate methodological counterparts. Network analysis, in its focus on relational aspects of data, can be used as an analytical approach that matches these theoretical developments and allows for their operationalisation. By reformulating current research questions in the language of network analysis, a range of methodological tools becomes accessible for the study of historical entanglements.

Secondly, the mathematical foundations of network analysis in graph theory provide a common methodological basis that spans disciplines. Contributions to the analysis of complex networks from biology or physics can advance analyses in history or philology, since the abstract representation of networks as graphs and their mathematical description as matrices is largely detached from the concrete phenomena they describe.[2] In the last years, network analysis has become more accessible for humanities scholars, thanks to the availability of powerful yet easy to use software tools like Gephi (Bastian, Heymann, Jacomy 2009) or Cytoscape (Shannon, Markiel, Ozier, Baliga, Wang, Ramage, Amin, Schwikowski, Ideker 2003).

This universality of the descriptive language is, however, different from the universality of the phenomena under study. While the seemingly universal applicability of network models might suggest a "natural" occurrence of network patterns (cf. prominently Albert, Barabasi 2002), the networks we study in the humanities are not discovered, but created. In order to obtain networks, scholars usually apply implicit or explicit rules of how to turn information from their sources into networks. This process of deriving networks as secondary data from primary sources requires careful planning and critical evaluation, not only because it provides the basis for all subsequent analyses, but also because it is not a neutral process: Assumptions about the character of relations in the data and theoretical knowledge inform these steps. In the humanities, data is often not collected in the form of network data, in contast e.g. to social network analysis in sociology, where data collection instruments exist for network data. The

[1] While this paper contains an analysis of this particular text, I want to stress that the focus of the paper is on methodology. Far from being an expert on Novalis or romanticism, I can only attempt to illustrate a potential use case for the method. Feedback from scholars more familiar with the material is most welcome.

[2] Of course, not all real-world networks share the same characteristics, so differences in the underlying network model can have significance, and some analytical measures might be more appropriate for some networks than for others. But still, all networks share the basic elements: Nodes and Edges.

network data under study in the humanities are often second-order representations that highlight one specific property of the data while neglecting others. In this sense, network analysis in the humanities is always the analysis of artefacts.[3]

2 How to read through networks

The prerequisite for applying network analysis is, of course, network data. Usually, humanities scholars work with texts. So in order to apply network analysis to the study of texts (or historical phenomena described in texts), an intermediate step of deriving network data from textual data is required. Approaches to generate networks from textual data vary greatly. They depend on the available sources, the research questions, and other factors. At the most fundamental level, it starts with the questions of what are the nodes, what are the relations, and what are the rules for their identification in the source material. Research on the subject has found different answers to these questions, but in general, approaches can be sorted into two main groups: networks *from* texts and networks *of* texts. In many applications of historical network research, the guiding principles are taken from social network analysis. Social actors, mainly individuals, are the nodes, social relations like marriage or fealty are the relations. Information about nodes and relations of interest is often collected from a variety of available sources and obtained through classical historical work (cf. e.g. Gramsch 2013). In archaeological research, where data is even sparser, nodes might also be settlements, and their relations inferred by geographical proximity (cf. e.g. Collar 2008). Since these studies are mainly interested in statements about a historical situation, this can be described as the historical approach: networks are built from texts, which themselves act as a source of information, not an object of study on its own right. The second main approach is concerned with the text as a literary work. Work in this area is concerned with building and analysing a network representation of texts. By considering words as nodes, and textual structures as edges, network analysis can be understood as a tool for re-reading texts. In this sense, network analysis becomes a technique of distant reading. I call this approach, where networks of texts are built, the literary (or philological) approach. In the remainder of this article, I will focus on this latter approach and introduce a new implementation.

Distant reading techniques, by their very definition, focus on large-scale structures and not on details. This is counter-intuitive for many philologists, and while there are good arguments for distant reading, certain research questions require greater care for detail. Network analysis, I want to argue, can be used as a bridge between the main advantages of distant reading and philologist's need for greater emphasis on the microstructure of texts. Novalis' essay will be used as an example to illustrate the point of network analysis as an instrument between close reading and distant reading. The metaphorical bridge can be seen as a two-way connection: From the close reading of detail to an overview of large-scale structure, and back to the examination of interesting details in the texts themselves.

2.1 Network Creation: From close to distant reading

Building networks from texts is, I argue, a process of construction rather than discovery. It cannot be assumed that there is one network that is the most appropriate representation for a text. Instead, the process of network creation requires a series of choices that depend on the research question and the theoretical and methodological background. This does not mean, however, that these created networks cannot be the starting point for discoveries. My argument is simply that the networks created from texts are a heuristical device that allows for a new perspective on the text, they are not the text or an inherent property of it. The two main questions when creating networks from texts are: What are the nodes? And what are the edges between them? Many distant reading methods use a "bag of words" approach: If words appear in a certain common context, then they are assumed to be related. This is known

[3] Cf. Giessmann (2006) for a cultural history of the concepts of nets and networks.

e.g. from topic modelling, but also applied in text network analysis (Paranyushkin 2011; Diesner, Carley 2005). For the purpose of network analysis, commonly a window of two sentences or five words is used. The pragmatic assumption of relatedness of cooccurring words often yields usable results, but is problematic in certain regards. Especially, this method is vulnerable to false positives, i.e. assuming relations where none actually are (Ferrer i Cancho, Solé, Köhler 2004, p. 1). This poses the question of which better criteria can be used to assume two words to be related. To enhance the quality of the network data, distant reading can learn from close reading.

For the identification of relations between words in a sentence, the inherent structure of language can be taken into account. Instead of using a bag of words model, the starting point for the proposed method is the syntactic structure of language. The rules for linking words are then derived from the rules of grammar. The dependency formalism (Tesnière 1980) provides a good starting point for network creation. The dependency grammar represents sentences as trees in which words are used as nodes, and dependents are positioned as child nodes of the word on which they depend (governor). In contrast to constituency grammars, the syntax trees contain only word nodes, not nodes that represent phrases. Ferrer i Cancho, Solé and Köhler (2004) have shown that this property allows to create word networks that represent units larger than a sentence, e.g. a text or a corpus. The syntactic dependency tree is treated as a proto-network, and by applying a set of transformation rules, a network representation for each sentence is created. These sentence networks are then merged to create the network for the whole text under investigation. Here, I follow the basic idea of the "syntactic dependency network" (Ferrer i Cancho, Solé, Köhler 2004, p. 2), but depart from the strict linguistic concept of dependency in favor of a more semantic approach. The individual transformation steps are described below. The process is illustrated in Figure 1, using this sample sentence[4]:

> Kindliches Zutrauen knüpfte die Menschen an ihre Verkündigungen.
>
> Childlike faith bound the people to their teachings.

[4] Cf. Novalis (1996) for the English translation

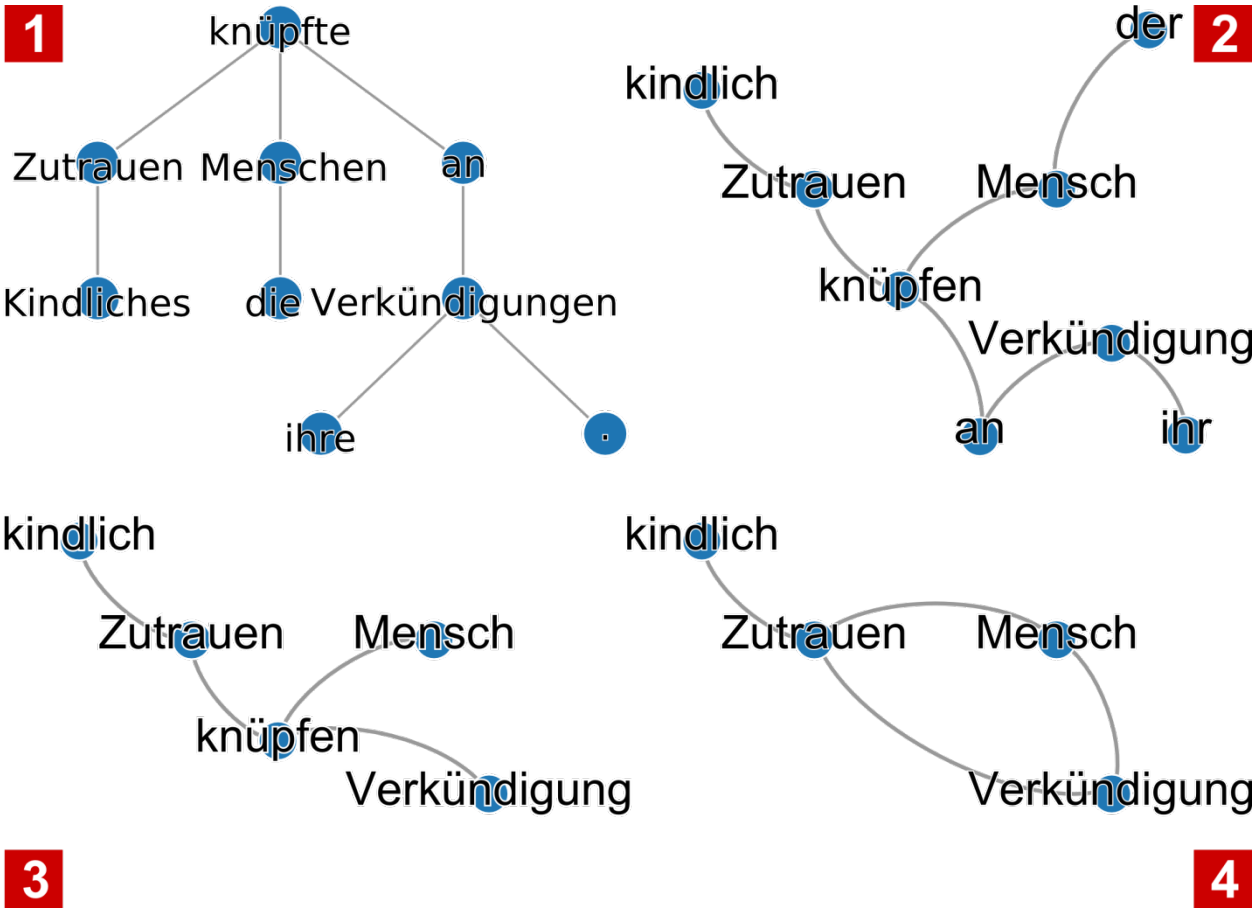

Figure 1: Transformation steps from dependency tree to network

In contrast to social networks, semantic networks do not only include social actors like persons or groups. Since the idea of semantic networks it to analyse meaningful relations between words, the network will include a much broader set of nodes. For a network representation of a text, the starting point for node identification are all words in the text. This is, however, neither semantically sensible nor manageable. In order to get to a more meaningful set of nodes, two techniques are typically applied: generalisation and filtering.

The aim of generalisation is to find a common generic concept for a set of semantically related word forms. This concept will then be used as a node in the network and represent all instances of words in the set. A relatively simple approach for this is stemming, i.e. reducing inflected word forms to a common root. Since words of different word classes can be reduced to a common root, the generalisation is rather broad and spans word classes. Depending on the application, this may or may not be desired behaviour. Using the Snowball stemmer for German (Porter 2001), the example sentence would be stemmed as “kindlich zutrau knupft die mensch an ihr verkund .” A more precise method is lemmatisation. Lemmatisation is the process of canonical form (or dictionary form) for a word. Since lemmatisation is dictionary based, it takes part of speech into account and will consequently result in a larger number of distinct nodes. Lemmatisation of the same sentence results in the sequence “kindlich Zutrauen knüpfen der Mensch an ihr Verkündigung .” In this application, lemmatisation has been used. Further generalisation might be achieved by using additional lexical resources to merge synonyms. Additionally, the result of the generalization process may still contain ambiguous entries, which may require further processing.

After generalisation, there is still a large number of nodes left, some of which carry no significant semantic value. The number of nodes can be reduced further by filtering out irrelevant lemmas. For filtering, often a list of "stop words" is used. These words are believed to carry only little meaning and are thus excluded from further analysis. However, the creation of a list of stop words requires significant work, and a single list of stop words might not be appropriate for all applications. Instead of listing all stop words explicitly, this method applies a filtering mechanism that is based on word classes. Closed word classes, like prepositions, conjunctions, or pronouns, are assumed to carry little or no semantic meaning. By using the result of part-of-speech tagging, all words that belong to a closed word class can be filtered out.

In addition, we assume that while verbs play an important role in relating words, they are less important as nodes. The proposed method drops verbs, but adds additional edges between the dependants of verbs, which were not originally part of the dependency tree (step 4 in Figure 1). As a further consequence of removing verbs, adverbs are also dropped. This leaves mainly nouns and adjectives as meaning carrying nodes. Again, all of these decisions can be taken differently. E.g., one can include verbs in the network, which will lead to a more linguistic and less semantic network.

2.2 Data and Tools

The digital edition of Novalis' essay has been taken from the TextGrid Repository[5]. It is a digitised version of the edition by Kluckhohn and Samuel (1968). The annotated TEI version is available under a Creative Commons Namensnennung 3.0 Deutschland license (CC-BY), but the text itself is in the public domain. The spelling was manually normalised in order to improve the automatic linguistic annotation. The annotation itself was carried out through the WebLicht service (Hinrichs, Hinrichs, Zastrow 2010), using the Stuttgart Dependency Parser toolchain for part-of-speech tagging, lemmatisation, and dependency parsing (Bohnet 2010).

From the annotated text, a network was created using a software developed by the author. The software is implemented as a WebLicht compatible service. It is available on GitHub under a GPL license.[6] The networks used in this example are created using this command:

```
annotators/dependency.py \
--nodes semantic --edges semantic \
< Novalis_Christenheit.tcf \
| exporters/graphml.py \
> Novalis_Christenheit.graphml
```

The analysis was performed using the igraph package (Csardi, Nepusz 2006) for the statistical environment R (R Development Core Team 2011). Network plots were created using Gephi (Bastian, Heymann, Jacomy 2009).

[5] http://textgridrep.de/browse.html?id=textgrid:sv8d.0
[6] https://github.com/SeNeReKo/TCFnetworks

2.3 Network Analysis: From distant to close reading

The network resulting from the described text-to-network method has 1137 nodes and 1748 edges. If one plots the full network, hardly anything will be visible. With a network of this size, the intuitive approach through visualisation is limited. In order to obtain information from this graph, mathematical measures can be used. Just like word frequencies can be used to get an impression of important topics in a text, centrality measures can be used to identify central nodes in a network. A node is central in a network if it is structurally important. There are different measures for centrality which have different meaning (cf. Scott 2000 for an overview). The interpretation of these measures in the context of text network analysis will also differ from applications in social network analysis. The simplest measure is the degree of a node, which is the number of nodes directly related to a given node. A concept is central in this regard if it is related to many other concepts. The concepts with the highest degree, as shown in table 1, give a first impression of the text content.

node	degree
Zeit	42
Mensch	38
neu	37
Kraft	22
heilig	22
groß	21
Welt	21
Glaube	21
Wissenschaft	21
alt	21

Table 1: The 10 most central concepts

While centrality might be an interesting concept that goes beyond simple frequency counting, it does not tell much about the internal structure of the whole network. By just listing concepts with a high degree, the structural information carried by lesser central nodes is lost. In order to get an appropriate "distant" view of the text, a

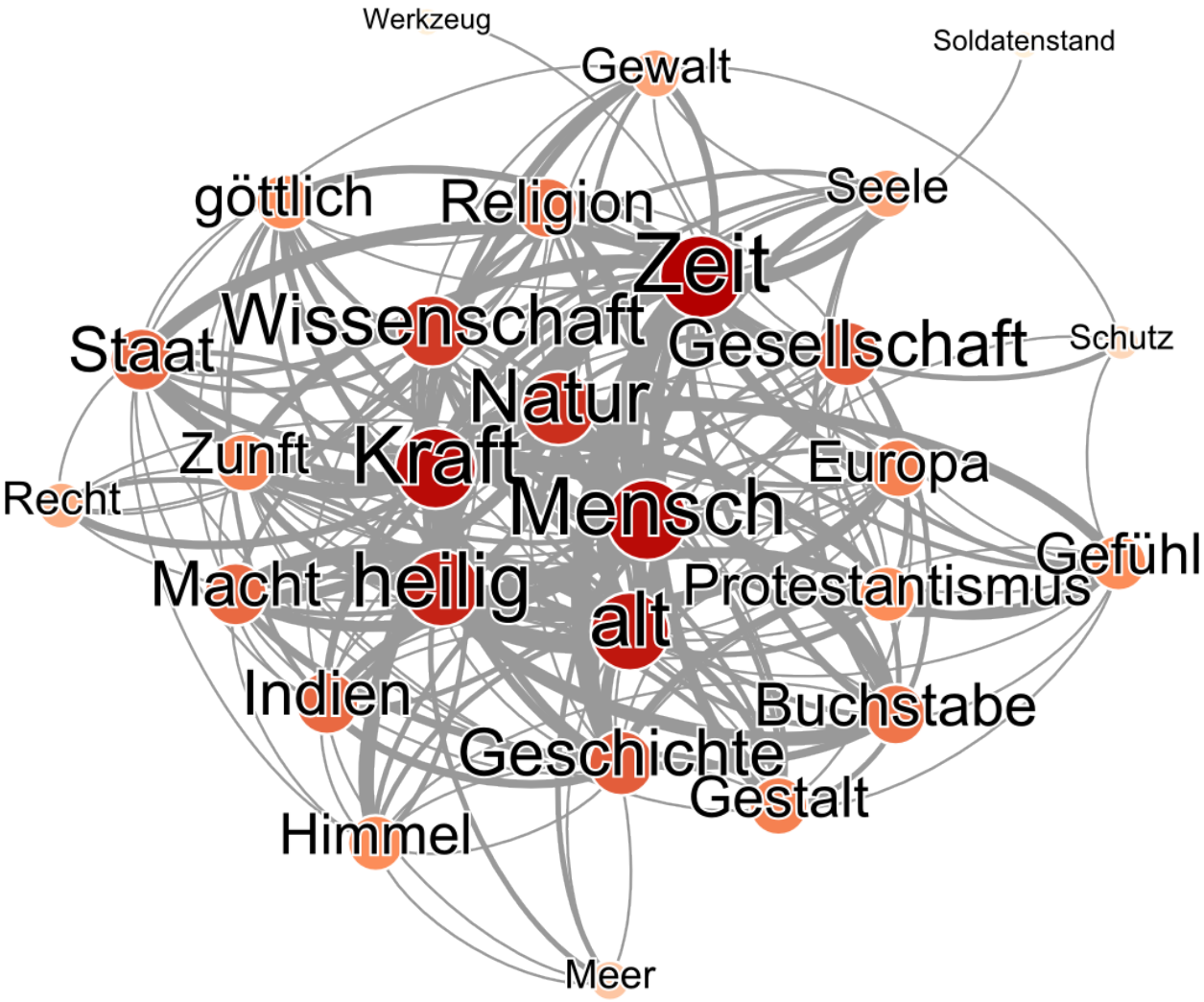

Figure 2: The community structure

summarisation method has to be applied that a) reduces noise compared to the full network, while b) keeping the

structural information of the complete network. For this purpose, the network topology can be used to find areas with greater density of relations, so-called communities (cf. Fortunato 2010 for an overview). Community detection is a clustering process that tries to find groups of nodes that contain more links between the group members than links to nodes outside the group. The community structure of the network gives information about groups of related concepts.[7] The division of a network into optimal communities is a computationally hard problem. Most current algorithms apply strategies that allow to find relatively good communities in reasonable time. In this paper, the multilevel algorithm by Blondel et al. (2008) has been used to find communities. This algorithm follows a bottom-up strategy: It successively merges nodes into larger communities, trying to improve the clustering in each step (Blondel, Guillaume, Lambiotte, Lefebvre 2008, p. 4). The text network contains 28 communities, with a minimum of 8 and a maximum of 89 members.

[7] The idea of related concepts bears some similarity with the idea of topics in topic modelling (e.g. Blei, Ng, Jordan 2003). However, the conceptual and mathematical foundations are largely different, as are the results.

These communities can be used for a second look at the text network. The degree measure of centrality, as listed above, identifies individual nodes by their structural properties with regard to the complete network. Using information about the community structure, i.e. about groups of closely related nodes, statements about central concepts can be refined. Instead of looking at global centrality measures, one can look at the centrality of concepts within their community. It has been argued that this gives us a more adequate view of the relative importance of nodes (Mutschke 2010). Additionally, the community structure can be used to infer the higher-level network structure by looking at how the communities themselves are related. Combining this information, the central concepts of the text can be visualised as a network of communities, using the most central concept of each community as its label, as shown in Figure 2.

Studying this network image can be seen as a practice in distant reading.[8] It highlights the most central word clusters of the text and visualises their position within the network structure. A striking result of this view is, for example, the centrality of the concept of time ("Zeit") for the text. And this is a plausible result, as it highlights the structure of Novalis' essay, which consists of images of the past (the idealised medieval times, the reformation), the present (the enlightenment) and a utopian future (in which Europe's Christian union is restored) (cf. Küng 1985). Similarily, other prominent nodes represent major topics of his time, e.g. science ("Wissenschaft") and religion ("Religion"), seen as opposing forces (cf. Sändig 1999, p. 349).

While this view highlights some important word clusters, it also hides a lot of analytically relevant detail. Central concepts, which just happen not to be the most central concept in any community, do not appear in this visualisation. This is also true for the word "Christenheit" (Christianity), which is, after all, part of the title. Thus, a closer look at the result of the analysis can reveal further insights. The following list shows the ten largest communities as a list of words, sorted from largest community to smallest, and within the community sorted from most to least central word:

1. Zeit, neu, Kirche, Genosse, sichtbar, schön, Glied, Glauben, Periode, Vertilgung
2. Wissenschaft, Hand, Feind, Bedürfnis, Geistlichkeit, Gelehrsamkeit, Art, Schatz, Bemühung, gelehrte
3. heilig, Geist, Menge, Christenheit, unendlich, Sinn, Gemüt, Geschäft, Musik, Belebung
4. Mensch, Leben, himmlisch, Handelsverkehr, Achtung, zutrauen, Rest, irdisch, Not, Behältnis
5. Buchstabe, Organ, gut, Kenntnis, Kopf, Traum, Erzeugnis, Verkündigung, Ohnmacht, fürchterlich
6. alt, Welt, Glaube, Verfassung, Wissen, Personalhaß, Schmuck, Himmelreich, Reiche, Anarchie
7. Kraft, groß, Teil, Interesse, Anlage, Geschlecht, Wachstum, weltlich, menschlich, Oberhaupt
8. Macht, Kind, Mutter, Andacht, geistlich, wohltätig, Gegenwart, Bruder, Keim, mannigfach
9. Religion, Freude, Gehorsam, Bild, vorzüglich, Inhalt, Entwurf, Name, Weise, unverkennbar
10. Geschichte, Stoff, merkwürdig, historisch, Menschheit, Krieg, Evolution, Versuch, letzter, Jahrhundert

The list representation provides a first impression of related words. The word "Christenheit" turns out to be part of the community which contains "heilig" (holy) as its most central node. This already gives an impression of the context of the word. But the network representation carries additional information about the internal structure of

[8] Since this analysis still deals with a single text, one might argue if this can already be called "distant reading". However, the same method can also be applied to a larger corpus.

each community. This allows to examine the organisation of word around central concepts more closely. Figure 3 shows the internal structure of this community. The visualisation shows how low-degree words group around the central nodes of the community.

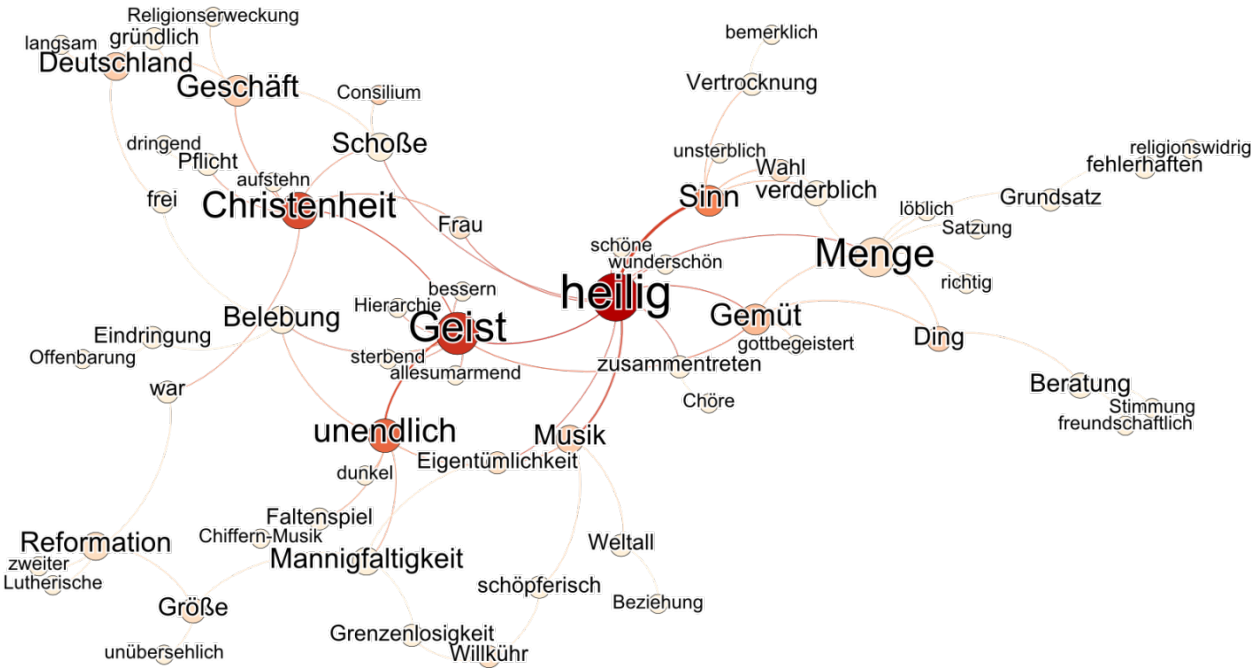

Figure 3: The "Christenheit" community

"Religion" is part of a different community (Figure 4), so it obviously has a different conceptual context than "Christenheit". This is an interesting result, as it counters the idea of Christianity being the only religion, or even Christianity and religion being equal, which was a prominent idea through much of Europe's history (cf. Daggers 2010). By examining the community strucutre, it is possible to find out more about the different ways Novalis conceptualises "Religion" in contrast to "Christenheit". "Christenheit" shares its community with words connected to aspects of transcedence, like "heilig" (holy), "Geist" (spirit), or "unendlich" (infinite) – concepts nowadays typically associated with religion. In Novalis use of the word, however, "Religion" is rather connected to words like "Freude" (joy), "Bild" (image), or "Entwurf" (blueprint). At the same time, this community contains no words connected to transcendence.

These initial findings evoke hypotheses which can serve as a starting point for further investigation. The surprisingly earthly context of "Religion" indicates that the main connotation of this term is not related to otherworldly matters, but rather to a specific aspect of human nature. This is supported by studies which show the close link between religion and the arts in Novalis' work (cf. Maatsch 2008). And there is also support in the network for this interpretation: The relation between "Religion" and any other node which occurs most frequent is the one with "Zeugungselement" (creative element). For a proper interpretation of this relation, close reading may

turn out more appropriate than the interpretation based on the distant view alone. Since the relations in this network are not abstract mathematical correlations, but refer to actual syntactical relations, it is possible to go back to the

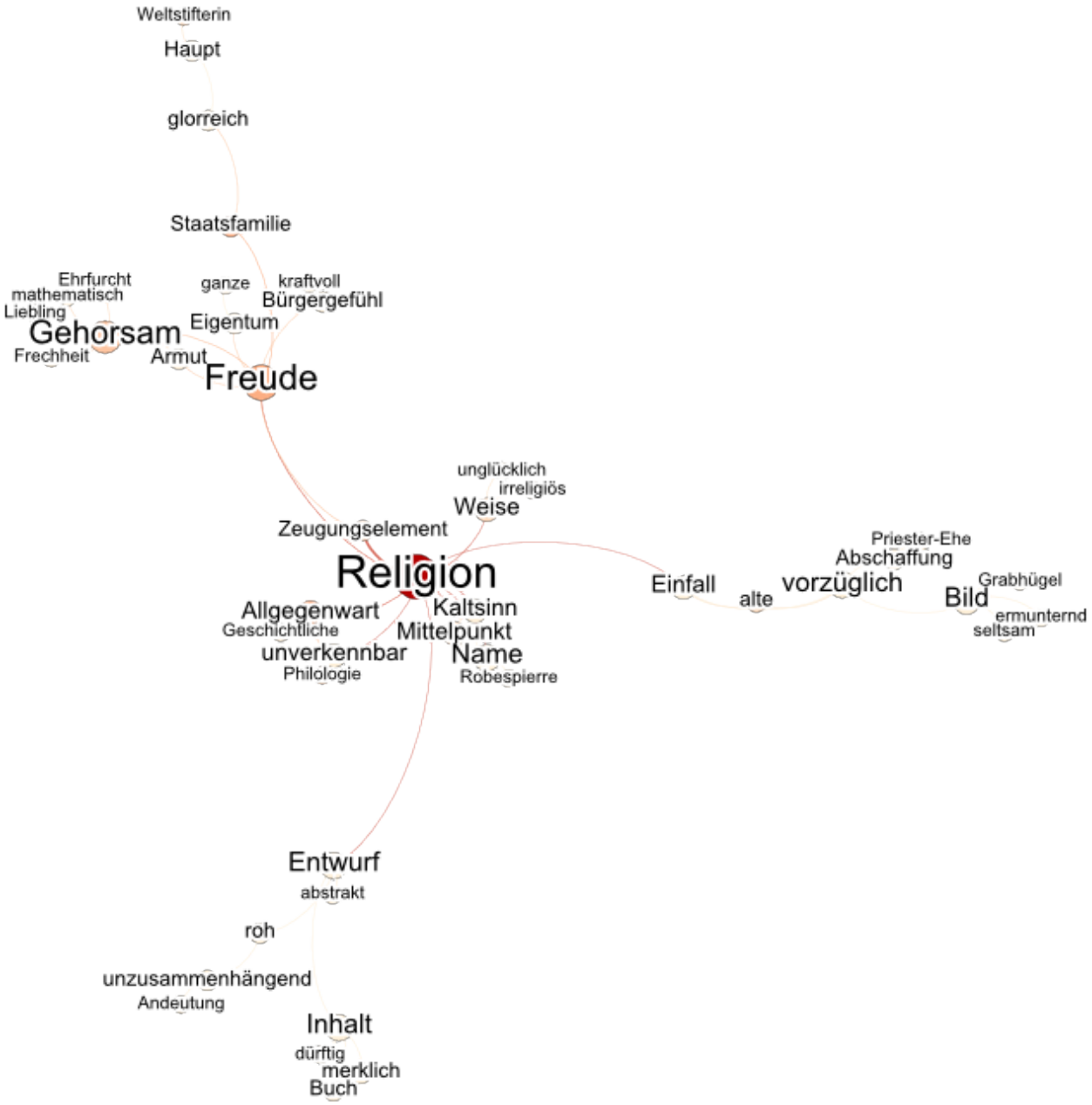

Figure 4: The "Religion" community

concrete source sections they originate from. During network creation, the program stores which words in the text are represented by which node, and which pairs of words are the basis for which edge in the network. This allows to link the network and the text directly, highlighting the corresponding source passages to a node or relation.

There are two sections which tell about this relation:

> Wahrhafte Anarchie ist das Zeugungselement der Religion. Aus der Vernichtung alles Positiven hebt sie ihr glorreiches Haupt als neue Weltstifterin empor.[9]

> Das Christenthum ist dreifacher Gestalt. Eine ist das Zeugungselement der Religion, als Freude an aller Religion. Eine das Mittlerthum überhaupt, als Glaube an die Allfähigkeit alles Irdischen, Wein und Brod des ewigen Lebens zu seyn. Eine der Glaube an Christus, seine Mutter und die Heiligen.[10]

A close reading of these sections, in turn, also tells us more about Novalis' conception of religion, Christianity, and their relation to each other. In the second section, Novalis describes religion as being one element of Christianity, as being a joyful, creational element. But Christianity as a historical phenomenon consists of more than this: It also provides a mode of relating the transcendent to the immanent sphere, and a specific set of beliefs. On the other hand, religion is free of these specifics that make Christian belief. On the contrary, as the first sections explicates: Religion gains its creative power once "everything positive" is destroyed. In this use of the word, "positive" refers to "everything established or laid down by authority" (Novalis 1996, p. 72 fn. 26). In consequence, the creative quality of religion can be seen as contrasting Christianity as positive religion. But Novalis does not see religion and Christianity as opposing forces, he integrates the "creative element" of religion into his conception of Christianity. In this sense, Novalis does not defend Christianity as the "absolute religion" (Daggers 2010, p. 967), like many of his contemporaries. Instead of putting Christianity into the centre of his conception of religion, he puts religion into the centre of his conception of Christianity, making the latter more than just one reification of the former.

3 Conclusion

This paper introduced a new method of network-based text analysis. The proposed method of generating text networks from dependency trees, as illustrated by the example, can serve as a link between close and distant reading. This connection works in both directions: Firstly, the distant view of the material is created by closely following the syntactical structure of the text. Thus, the distant reading can be seen as an abstraction of a form of close reading. Secondly, results of distant analysis can directly be linked back to source passages. This allows for evaluation of the results of network analysis, but also provides starting points for more hermenutical approaches to interpretation. This allows for a division of work between distant and close reading: Distant reading can provide the ground for new discoveries and interesting hypotheses, while close reading serves as a means to test those hypotheses with the source material.

Research on the described network extraction methodology is still in progress. Further testing and analysis is expected to result in improvements on the exact procedure. In addition, a thorough comparision with the results of competing methods, especially cooccurrence networks (cf. e.g. Biemann, Bordag, Heyer, Quasthoff, Wolff 2004) and topic models (Blei, Ng, Jordan 2003) still has to be done.

---

[9] True anarchy is the creative element of religion. From the destruction of everything positive it lifts up its glorious head as the creator of a new world.

[10] Christianity has three forms. One is the creative element of religion, the joy in all religion. Another is mediation in general, the belief in the capacity of everything earthly to be the wine and bread of eternal life. Yet a third is the belief in Christ, his mother and the saints.

References

ALBERT, Reka and BARABASI, Albert-Laszlo, 2002, Statistical mechanics of complex networks. *Reviews of Modern Physics* [online]. January 2002. Vol. 74, no. 1, p. 47–97. [Accessed 5 January 2014]. DOI 10.1103/RevModPhys.74.47. Available from: http://arxiv.org/abs/cond-mat/0106096

BASTIAN, Mathieu, HEYMANN, Sebastien and JACOMY, Mathieu, 2009, Gephi: an open source software for exploring and manipulating networks. In : International AAAI Conference on Weblogs and Social Media [online]. The AAAI Press. 2009. p. 361–362. [Accessed 6 March 2015]. ICWSM. Available from: http://aaai.org/ocs/index.php/ICWSM/09/paper/view/154

BERLIN-BRANDENBURG ACADEMY OF SCIENCES AND HUMANITIES, 2014, Thesaurus Linguae Aegyptiae. [online]. 31 October 2014. Available from: http://aaew.bbaw.de/tla/

BIEMANN, Christian, BORDAG, Stefan, HEYER, Gerhard, QUASTHOFF, Uwe and WOLFF, Christian, 2004, Language-Independent Methods for Compiling Monolingual Lexical Data. In : Computational Linguistics and Intelligent Text Processing [online]. Berlin, Heidelberg : Springer. p. 217–228. Lecture Notes in Computer Science, 2945. [Accessed 6 March 2015]. Available from: http://link.springer.com/chapter/10.1007/978-3-540-24630-5_27

BLEI, David M., NG, Andrew Y. and JORDAN, Michael I., 2003, Latent Dirichlet Allocation. Journal of Machine Learning Research [online]. März 2003. Vol. 3, p. 993–1022. [Accessed 6 March 2015]. Available from: http://dl.acm.org/citation.cfm?id=944919.944937

BLONDEL, Vincent D., GUILLAUME, Jean-Loup, LAMBIOTTE, Renaud and LEFEBVRE, Etienne, 2008, Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment [online]. 9 October 2008. Vol. 2008, no. 10, p. P10008. [Accessed 2 December 2013]. DOI 10.1088/1742-5468/2008/10/P10008. Available from: http://arxiv.org/abs/0803.0476

BOHNET, Bernd, 2010, Very high accuracy and fast dependency parsing is not a contradiction. In : Proceedings of the 23rd International Conference on Computational Linguistics [online]. Association for Computational Linguistics. 2010. p. 89–97. [Accessed 21 January 2014]. Available from: http://dl.acm.org/citation.cfm?id=1873792

BRAUNGART, Wolfgang, 2011, Subjekt Europa, Europas Subjekt. Novalis' katholische Provokation "Die Christenheit oder Europa." Sinn und Form. Beiträge zur Literatur. 2011. Vol. 63, no. 4, p. 544–558.

COLLAR, Anna, 2008, Networks and religious innovation in the Roman Empire. University of Exeter.

CONRAD, Sebastian and RANDERIA, Shalini, 2002, Einleitung : Geteilte Geschichten – Europa in einer postkolonialen Welt. In : Jenseits des Eurozentrismus: postkoloniale Perspektiven in den Geschichts- und Kulturwissenschaften. Frankfurt : Campus. p. 9–49.

CSARDI, Gabor and NEPUSZ, Tamas, 2006, The igraph software package for complex network research. InterJournal [online]. 2006. Vol. Complex Systems, p. 1695. Available from: http://igraph.org/

DAGGERS, Jenny, 2010, Thinking "Religion": The Christian Past and Interreligious Future of Religious Studies and Theology. Journal of the American Academy of Religion [online]. 1 December 2010. Vol. 78, no. 4, p. 961–990. [Accessed 30 January 2014]. DOI 10.1093/jaarel/lfq078. Available from: http://jaar.oxfordjournals.org/content/78/4/961

DIESNER, Jana and CARLEY, Kathleen M., 2005, Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis. Causal mapping for information systems and technology research: Approaches, advances, and illustrations [online]. 2005. P. 81–108. [Accessed 8 September 2013]. Available from: http://andrew.cmu.edu/user/jdiesner/publications/DiesnerCarley_meta_matrix_text_analysis.pdf

EMIRBAYER, Mustafa, 1997, Manifesto for a relational sociology. American Journal of Sociology [online]. 1997. Vol. 103, no. 2, p. 281–317. [Accessed 29 October 2013]. Available from: http://www.jstor.org/stable/10.1086/231209

FERRER I CANCHO, Ramon, SOLÉ, Ricard V. and KÖHLER, Reinhard, 2004, Patterns in syntactic dependency networks. Physical Review E [online]. Mai 2004. Vol. 69, no. 5, p. 051915. [Accessed 21 November 2012]. DOI 10.1103/PhysRevE.69.051915. Available from: http://link.aps.org/doi/10.1103/PhysRevE.69.051915

FORTUNATO, Santo, 2010, Community detection in graphs. Physics Reports [online]. February 2010. Vol. 486, no. 3–5, p. 75–174. [Accessed 23 October 2013]. DOI 10.1016/j.physrep.2009.11.002. Available from: http://www.sciencedirect.com/science/article/pii/S0370157309002841

GIESSMANN, Sebastian, 2006, Netze und Netzwerke: Archäologie einer Kulturtechnik, 1740-1840. Bielefeld : Transcript.

GRAMSCH, Robert, 2013, Das Reich als Netzwerk der Fürsten: politische Strukturen unter dem Doppelkönigtum Friedrichs II. und Heinrichs (VII.) 1225-1235.

HINRICHS, Erhard W., HINRICHS, Marie and ZASTROW, Thomas, 2010, WebLicht : Web-Based LRT Services for German. In : Proceedings of the ACL 2010 System Demonstrations [online]. 2010. p. 25–29. [Accessed 6 March 2015]. Available from: http://www.aclweb.org/anthology/P10-4005

KRECH, Volkhard, 2012, Dynamics in the History of Religions. Preliminary Considerations on Aspects of a Research Programme. In : Dynamics in the history of religions between Asia and Europe in past and present times. Leiden : Brill. p. 15–70. Dynamics in the history of religions.

KÜNG, Hans, 1985, Religion im Spiegel romantischer Poesie. In : Dichtung und Religion. Pascal, Gryphius, Lessing, Hölderlin, Novalis, Kierkegaard, Dostojewski, Kafka. München : Kindler. p. 164–182.

MAATSCH, Jonas, 2008, "… wenn nicht eine Anziehung gegen den Himmel sie auf der Höhe schwebend erhält." : Novalis' "Europa"-Text und die Kraft der Poesie. In : Europa in Weimar: Visionen eines Kontinents [online]. Göttingen : Wallstein. p. 239–255. [Accessed 11 October 2013]. Available from: http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/14314

MORETTI, Franco, 2013, Distant Reading. London : Verso.

MUTSCHKE, Peter, 2010, Zentralitätsanomalien und Netzwerkstruktur. Ein Plädoyer für einen „engeren" Netzwerkbegriff und ein community-orientiertes Zentralitätsmodell. In : Netzwerkanalyse und Netzwerktheorie [online]. VS Verlag für Sozialwissenschaften. p. 261–272. [Accessed 23 October 2013]. Available from: http://link.springer.com/chapter/10.1007/978-3-531-92029-0_20

MÜTZEL, Sophie and FUHSE, Jan, 2010, Einleitung: Zur relationalen Soziologie Grundgedanken, Entwicklungslinien und transatlantische Brückenschläge. In : Relationale Soziologie [online]. VS Verlag für

Sozialwissenschaften. p. 7–35. [Accessed 16 January 2015]. Available from: http://link.springer.com/chapter/10.1007/978-3-531-92402-1_1

NOVALIS, 1968, Die Christenheit oder Europa. Ein Fragment. In : Novalis. Schriften. Die Werke Friedrich von Hardenbergs. Zweite, nach den Handschriften ergänzte, erweiterte und verbesserte Auflage. Darmstadt : Wissenschaftliche Buchgesellschaft. p. 507–527.

NOVALIS, 1996, Christianity or Europe: A Fragment. In : The Early Political Writings of The German Romantics. Cambridge University Press. p. 59–80. Cambridge Texts in the History of Political Thought.

OSTERHAMMEL, Jürgen, 2001, Geschichtswissenschaft jenseits des Nationalstaats: Studien zu Beziehungsgeschichte und Zivilisationsvergleich. Göttingen : Vandenhoeck & Ruprecht. Kritische Studien zur Geschichtswissenschaft, 147.

PARANYUSHKIN, Dmitry, 2011, Identifying the Pathways for Meaning Circulation using Text Network Analysis [online]. Nodus Labs. [Accessed 6 March 2015]. Available from: http://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis/

PORTER, Martin F., 2001, Snowball: A language for stemming algorithms. [online]. 2001. [Accessed 5 March 2015]. Available from: http://snowball.tartarus.org/texts/introduction.html

R DEVELOPMENT CORE TEAM, 2011, R : A Language and Environment for Statistical Computing [online]. Vienna, Austria : R Foundation for Statistical Computing. Available from: http://www.R-project.org/

SÄNDIG, Brigitte, 1999, Was kann Religion? Chateaubriands Génie du christianisme und Novalis' Die Christenheit oder Europa. Romanische Forschungen : Vierteljahrsschrift für romanische Sprachen und Literaturen. 1999. No. 3, p. 336–354.

SCOTT, John, 2000, Social network analysis: A handbook. London : Sage Publications.

SHANNON, Paul, MARKIEL, Andrew, OZIER, Owen, BALIGA, Nitin S, WANG, Jonathan T, RAMAGE, Daniel, AMIN, Nada, SCHWIKOWSKI, Benno and IDEKER, Trey, 2003, Cytoscape : a software environment for integrated models of biomolecular interaction networks. Genome research. November 2003. Vol. 13, no. 11, p. 2498–2504. DOI 10.1101/gr.1239303.

TESNIÈRE, Lucien, 1980, Grundzüge der strukturalen Syntax. Stuttgart : Klett-Cotta.

WERNER, Michael and ZIMMERMANN, Bénédicte, 2003, Penser l'histoire croisée : entre empirie et réflexivité. Annales [online]. 2003. Vol. 58, no. 1, p. 7–36. [Accessed 5 January 2014]. Available from: http://www.cairn.info/article.php?ID_ARTICLE=ANNA_581_0007

Biography of the author

Dr. **Frederik Elwert** has studied religion and sociology in Bremen and Leipzig. He is currently post-doctoral researcher at the Center for Religious Studies (CERES) in Bochum and project manager of "SeNeReKo", a joint research project of CERES and the Trier Center for Digital Humanities. He works on network analysis methodology and the interplay of religion and migration.

Acknowledgements

The SeNeReKo project upon which this publication is based is funded by the German Federal Ministry of Education and Research under the project number 01UG1242A. The author of this paper is responsible for its content.