A Distant Reading Visualization for Variant Graphs

Stefan Jänicke, Image and Signal Processing Group, Leipzig University, Germany
Annette Geßner, Goettingen Centre for Digital Humanities, University of Göttingen, Germany
Gerik Scheuermann, Image and Signal Processing Group, Leipzig University, Germany

1. Motivation

A Variant Graph is a data structure that can be used to model the similarities and differences among various editions of a text (Schmidt et al., 2009). Last year we presented a set of design rules for Variant Graphs (Jänicke et al., 2014) that are implemented by the web-based tool TRAViz.[1] Figure 1 shows a Variant Graph for 24 different English translations of *Genesis 1:1*.
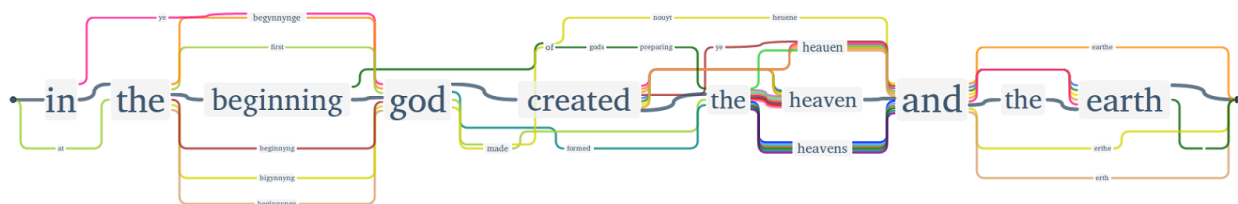

Figure 1: *Genesis 1:1* in 24 English translations.

Designing the graph that way, the humanist is able to analyze both similarities and differences among various text editions on verse level. Although TRAViz can also be used for larger text entities such as sections or chapters, the resultant visualizations are hardly readable and an analysis of the Variant Graphs becomes a laborious task. Therefore, TRAViz remains a close reading visualization tool only to be used when the user has a specific research question for a desired text part, such as the analysis of numerous different translations of the sixth commandment, commonly recited as "You shall not kill" (Jänicke et al., 2015). Research questions like "For which Bible books the various translations are most similar?" or "For which chapters of book X the rather similar translations A, B and C differ most and why?" cannot be answered with TRAViz. As Moretti suggests, such kind of questions require distant reading approaches to be answered (Moretti, 2005). This paper proposes a visualization that offers a distant view on Variant Graphs calculated for each Bible verse in order to support the humanists in detecting and analyzing occurring patterns on higher text hierarchy levels (the entire Bible, books, chapters) as provided by TRAViz (verses).

2. The Bible Corpus

The Bible is a very good use-case for the purpose of designing a distant reading visualization for Variant Graphs. Firstly, it is a very influential and well-known

text, which supports easy evaluation of results. Secondly, it's structure includes a four-level hierarchy that makes views of varying distance on the text possible: the Bible (level 1) consists of books, each book (level 2) is divided into chapters, and each chapter (level 3) is composed of verses. Each verse (level 4) can be visualized using a close reading visualization like TRAViz, but all other hierarchy levels require a distant reading solution.
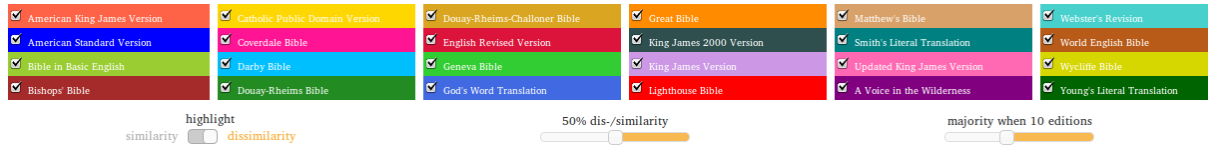
The Bible corpus of our project consists of 24 English translations spanning a period from the 14th (Wycliffe Bible) to the 21th century (e.g., Catholic Public Domain Version). It includes editions in Middle and Modern English, a great variety of translation dependencies (e.g., editions based on the King James Version) as well as a diversity of translation motivations (e.g., simplified language in the Bible in Basic English). The versatility of the corpus allows for a multitude of research questions to be asked by the humanists of our project. One of their use cases is outlined in Section 4.

3. Visualization Design

Figure 2 shows a screenshot of the system our humanists work with. Arranged in columns, the top panel lists all available Bible editions either sorted by year of publication or in alphabetical order. On demand, the user is able to compose a desired set of editions by clicking the corresponding checkboxes. Below the Bible editions, the humanist is able to tweak the visualization dependent on the given research question. In particular, the user can:
  (a) define a threshold value for the majority of editions,
  (b) either highlight text passages that are similar or dissimilar to a certain extent from the predefined majority, and
  (c) adjust the percentage of words that need to be similar or dissimilar regarding the majority to highlight a text passage.

The bottom panel visualizes a "(dis-)similarity fingerprint" for all 24 editions of the Bible on level 1; the user can interactively navigate between the different hierarchy levels. Each Bible book receives a rectangular block with its width reflecting the number of chapters. According to the configuration in this example, a tiny rectangle for a chapter of an edition is drawn in its assigned color,[2] if at least 50% of its contained words differ from the majority of at least ten editions. The resultant pattern reflects, e.g., three salient editions for the New Testament, which reveals individual translation styles: the Wycliffe Bible (1380, dark yellow) is the only translation in Middle English, whereas the God's Word Translation (1995, blue) and the Bible in Basic English (1949, light green) aim to be understood very easily nowadays, and thus, choose to deviate from other editions that tend to be more antiquated and sophisticated in language and style.
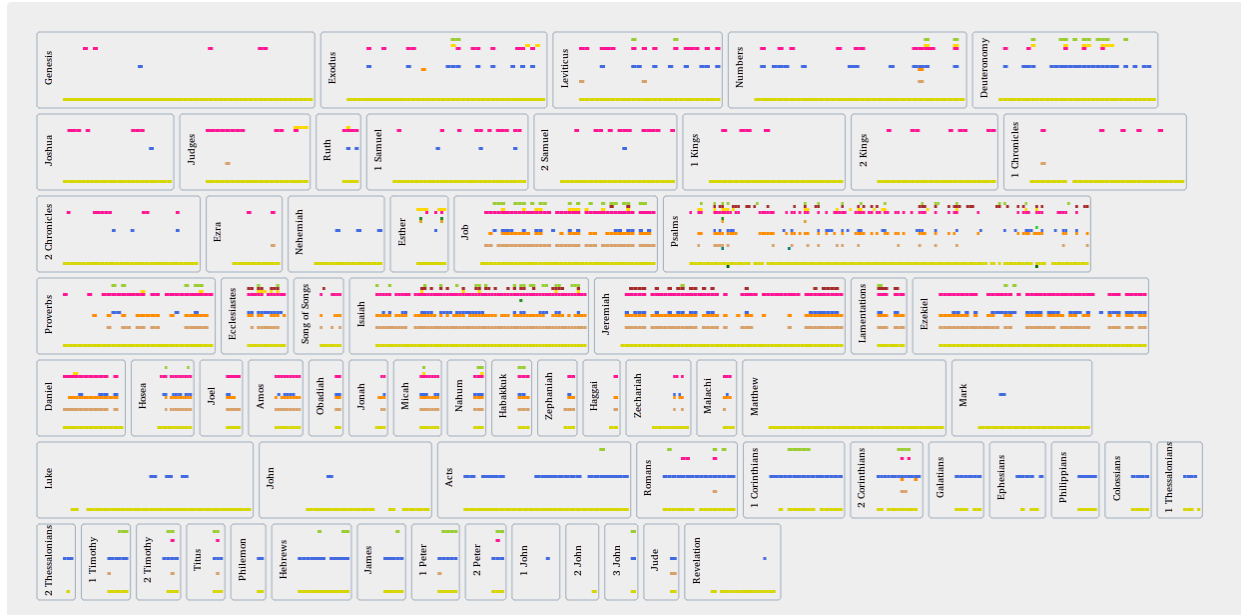
Figure 2: Distant reading of 24 Bible editions.

4. A Use Case

Concentrating on similarity (80% similarity, majority of seven) highlights Bible editions that are very similar in almost all chapters of all books (Fig. 3). Most of these editions are based on the King James Version, which is known as the most influential translation (Ryken, 2011). But surprisingly, albeit claiming to be "as exact a translation as possible"[3] in Modern English from the original languages Hebrew and Greek, the Darby Bible (1890, light blue) is also highlighted as very similar.

To determine the role of the Darby Bible, we remove all Bible editions not (apparently) connected to the King James Version. Now, clicking the book *Matthew* and highlighting dissimilarity (50% dissimilarity, majority of seven), we detect a cluster of derivations for the verses 16–18 of chapter 7 (Fig. 4). As implied on the book-level, the chapter-level confirms the Darby-deviations by highlighting the corresponding passages in the individual verses based upon a majority of four (Fig. 5).

Figure 3: The Bible (level 1) visualizing similarity.
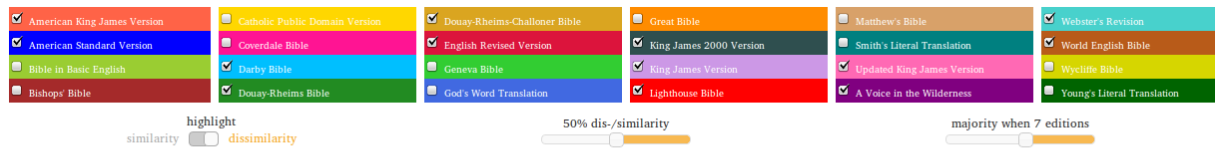
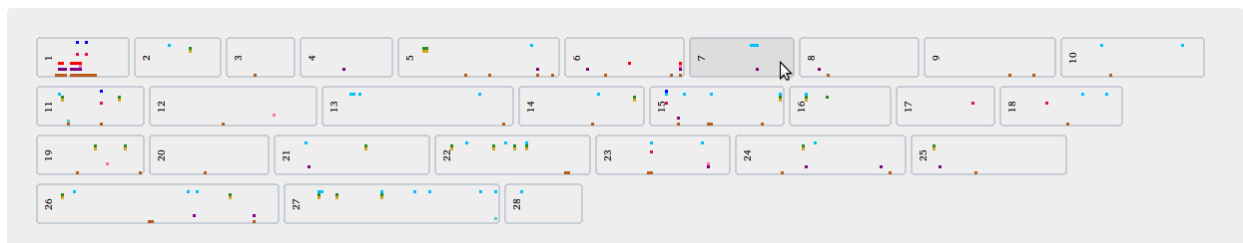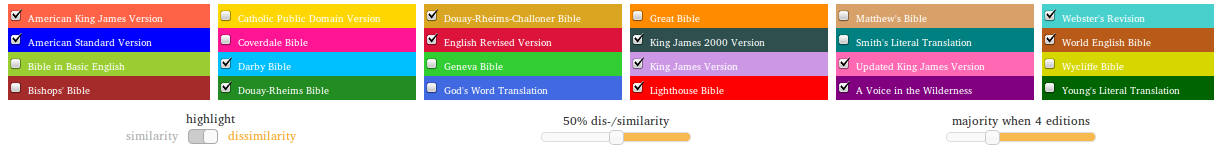**Figure 4:** Dissimilarity of selected editions in *Matthew* (level 2).

Figure 5: Dissimilarities of selected editions in *Matthew 7* (level 3).

Indeed in *Matthew 7:16* (Fig. 6), Darby and some other editions differ in word order (but not so much in the translations); among others he choses the elder word "ye" instead of "you" and as the only translation he writes "a bunch of grapes" instead of "grapes" and instead of "figs of thistles" he writes "thistles figs?". In *Matthew 7:17* (Fig. 7) and *Matthew 7:18* (Fig. 8) the main structure of the verse remains in Darby, but some words are translated differently: instead of "bring forth" Darby uses the word "produce" (followed by World English Bible [2000, brown] and A Voice in the Wilderness [2004, purple]), the word "nor" instead of "neither" (followed by A Voice in the Wilderness), and when fruit are described, Darby uses the not morally associated adjective "bad" instead of "evil", and "worthless" instead of "corrupt".
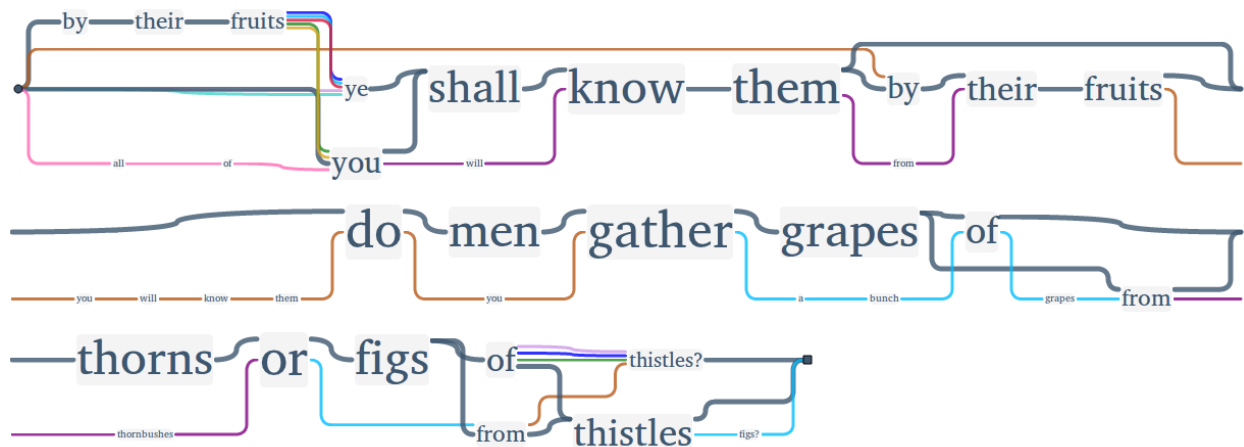
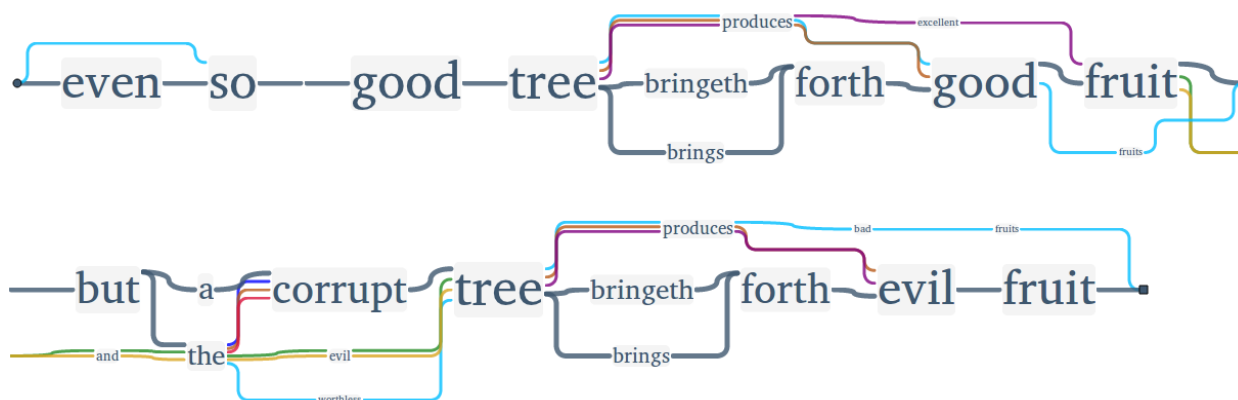

Figure 6: *Matthew 7:16* (level 4).
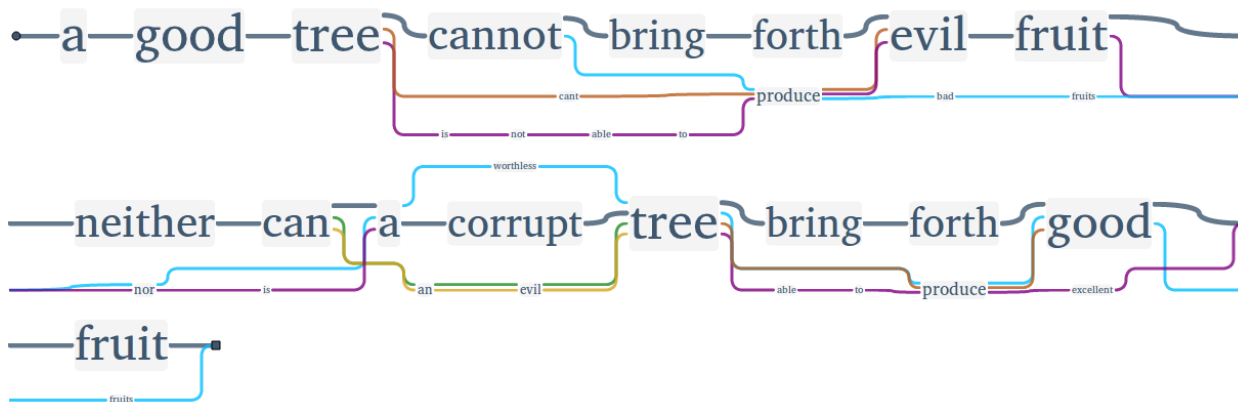
Figure 7: *Matthew 7:17 (level 4).*

Figure 8: *Matthew 7:18 (level 4).*

All in all, with this setting even when choosing passages marked as dissimilar, the close view confirms what the distant view implied, which is, how close the Darby Bible sticks to the other editions concerning word order, language and most of the words themselves. Most deviations are single words that are substituted by synonyms or something similar.

Nevertheless, the Darby variations tend to be the very obvious ones, next to those of A Voice in the Wilderness, but rarely seem to change the meaning of the text in a significant way. Thus, it seems that the Darby Bible, which tried to translate the ancient languages as exactly as possible, may have had a significant impact on later translations based (not only) on the King James Version.

5. Conclusion

In contrast to projects dealing with a rather small number of text editions on a close reading basis,[4] the presented novel technique provides a distant reading of Variant Graphs for a potential high number of editions. Just reading 24 translations

of the Bible would require a huge amount of time, and comparing those editions would take even longer. Being able to see similarity and dissimilarity on various text hierarchy levels enables the user not only to save time by directing to passages that could be of interest, but it can even raise new worthy research questions. This way of visualizing bears even the potential to cause serendipity by showing relations that have not been seen yet.

In the future, we plan to provide an Open Source library to enable the application of the visualization for other hierarchically structured texts. Visualizing the (dis-)similarity between various editions of Homer's epics could be one of the interesting examples. Being an important work for philologists, we could also extend the visualization to display the occurring transposed verses.

Acknowledgements

References

Harrower, M. and Brewer, C. A. (2003). ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal*, 40(1), 27-37.

Jänicke, S., Geßner, A., Büchler, M. and Scheuermann, G. (2014). 5 Design Rules for Visualizing Text Variant Graphs. In *Proceedings of the Digital Humanities 2014*.

Jänicke, S., Geßner, A., Franzini, G., Terras, M., Mahony, S., and Scheuermann, G. (2015). TRAViz: A Visualization for Variant Graphs. To appear in *Digital Scholarship in the Humanities*, 2015.

Moretti, F. (2005) Graphs, Maps, Trees: Abstract Models for a Literary History. *Verso*.

Ryken, L. (2011). The Legacy of the King James Bible: Celebrating 400 Years of the Most Influential English Translation. *Crossway*.

Schmidt, D. and Colomb, R. (2009). A data structure for representing multi-version texts online. *International Journal of Human-Computer Studies*, 67(6), 497-514.

Notes

1. http://www.traviz.vizcovery.org/
2. The fixed edition order simplified the crucial task of selecting colors. Following the suggestions made by Harrower (Harrower et al., 2003), we defined a set of 24 varying colors ordered by alternating hue and saturation intensity, so that visually similar colors are not placed next to each other.
3. According to the introduction to the 1890 edition, which can be found online at: http://www.ccel.org/bible/jnd/darby.htm#2
4. Other projects also provide solutions for the visualization of textual variance. CollateX (http://collatex.net/), primarily focused on developing alignment algorithms for text editions, uses the GraphViz library (http://www.graphviz.org/) to visualize rather small texts analogous to TRAViz. Some web-based platforms allow to read parallel texts in the browser. Juxta Commons (http://juxtacommons.org/) shows variant patterns between two texts and with Versioning Machine (http://v-machine.org/) multiple editions are displayed next to each other. Both tools support only close reading for a small number of text editions.