CohExplore: Visually Supporting Students in Exploring Text Cohesion

C. Liebers\textsuperscript{1}, S. Agarwal\textsuperscript{2}, and F. Beck\textsuperscript{2}

\textsuperscript{1}University of Duisburg-Essen, Germany
\textsuperscript{2}University of Bamberg, Germany

\begin{abstract}
A cohesive text allows readers to follow the described ideas and events. Exploring cohesion in text might aid students enhancing their academic writing. We introduce CohExplore, which promotes exploring and reflecting on cohesion of a given text by visualizing computed cohesion-related metrics on an overview and detailed level. Detected topics are color-coded, semantic similarity is shown via lines, while connectives and co-references in a paragraph are encoded using text decoration. Demonstrating the system, we share insights about a student-authored text.
\end{abstract}

1. Introduction

Students with limited experience and skills in writing face challenges in authoring a text that is easy to follow for the readers — the text might lack cohesion. Text cohesion is a property of the text that involves features that guide readers in interpreting substantial ideas \cite{GMLC04,GMK11,GM11}. Topics reflect the overall text structure and semantics. Semantic similarity, based on topic detection, is significantly higher for high-cohesion than low-cohesion texts \cite{MLMG10} and can be calculated using Latent Se-

\begin{figure}[ht]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{CohExplore shows (A) a colored bar for detected topics, (B) the document structure with similarity lines between (B\textsubscript{1}) the text headings and (B\textsubscript{2}) paragraphs, (C) the raw text with decorations and similarity arcs, and (D) an interactive legend.}
\end{figure}
We present CohExplore (Fig. 1), an approach that visualizes cohesion-related metrics to aid reflection. It consists of an overview showing similarities of chapters and paragraphs and a reading view. It does not judge the text quality or scores but invites exploring text features that contribute to its cohesion. Coh Explore aims to support analyzing texts by visualizing: (1) semantic text similarity, (2) connectives, and (3) co-references, covering global (1) and local (2, 3) features of text cohesion.

Regarding global cohesion, we use Latent Semantic Analysis (LSA) [LMDK07, Dos26] with a computed topic number [Nav09] to identify topics in a text by grouping frequently co-occurring words. CohExplore leverages these topics to compute the semantic similarity between text segments: pairs of sentences, paragraphs, and chapters. We compute the cosine similarity of their vectors, showing similarities of chapters and paragraphs and a reading view. Varying pie chart colors indicate a similarity of 0.2 in the shown paragraph (Fig. 2), indicating abrupt topic shifts (l. 263), and i. 264, and l. 265, and the absence of co-references might make the section harder to follow.

Figure 2: Selecting a paragraph shows the topic words in their respective color. Hovering the connectors tile in the legend highlights connectives through borders.

Document Overview The Document Overview on the left (Fig. 1B) is split into two columns and was introduced by VarifocalReader [KJW14]. The headings column (Fig. 1B1) shows the hierarchical structure of the text via indentation. In the paragraphs column (Fig. 1B2), colored stacked bars show the topic distribution of paragraphs. Users can scroll or click on headings, or drag a frame to navigate the document. The other columns follow. Additional lines depict semantic similarity among chapters, subchapters, and paragraphs, to support analyzing if similarities are expected or desired, ordered hierarchically to minimize overlap. Their opacity encodes their similarity.

Text Cohesion in Paragraphs To show aspects of local cohesion, the text panel (Fig. 1C, Fig. 2) highlights cohesion between sentences. Pie charts show topic distribution of sentences to spot unusual patterns like multiple discussed topics. Arcs depict the semantic similarity within a paragraph, their opacity encodes the similarity value. We chose slightly different visualizations for topic distribution and semantic similarity to visually discern them from the overview, which has more text and topic words. Text decorations display local features: co-references, connectives, and a word’s topic. Co-references, pointing to the previous or following sentence, are detected using word lemmas (stem), and are underlined in the first and given a gray background on subsequent occurrences. Commonly used connectives, indicating relations between arguments, are emphasized in bold to not interfere with the co-references encoding. When selecting a paragraph, words are colored and marked with superscript numbers of their topic (Fig. 2).

Legend and Interactions The legend on the right (Fig. 1D) explains cohesion features and visual encodings. Hovering highlights the feature, like emphasizing connectives with borders (Fig. 2). Clicking a bar or a pie chart reveals links connecting the corresponding segment, while hiding unrelated lines and arcs. A slider enables setting a minimal threshold for similarity links.

4. Conclusion and Future Work

We introduced a visualization approach that enables exploring text cohesion. It uses different cohesion metrics such as semantic similarity, connectives, and co-references. Future work includes adding co-reference cohesion metrics [MLMG10]; incorporating measures such as syntactic complexity, readability, and style word usage in its visualizations [GMLC04]; and applying alternative topic detection methods [ASFS18, BZSA18] to improve the topic quality, reduce their overlap, and refine the visualized text cohesion.

Acknowledgments

We thank Andrea Horbach and Torsten Zesch for valuable feedback on the approach.
References


