



Goethe-Universität Frankfurt
Institut für England- und Amerikastudien

and

Universität Konstanz
Fachbereich Sprachwissenschaft

VisaS Manual

Version: Nov. 4, 2019

Authors: Iverina Ivanova, Janina Radó

<http://visas.lingvis.io>

Introduction

The goal of [VisaS \(Visualisierung akademischer Schreibkompetenzen\)](#) is to provide a tool for writing tutors and instructors that enables them to automatically detect and visualize certain features of student essays. The visualizations are intended as a helpful instrument in analyzing students' writing skills in advising sessions and in the classroom. The tool allows a comparison of the text at hand with a cohort chosen by the user (papers by the same student, essays by classmates etc.)

With VisaS you can extract information from several texts and connect them in different modes of visualization. Currently the tool can analyze essays written in English. The dimensions of analysis are linguistic sophistication (e.g. the usage of complex sentences and specialized vocabulary), argumentation (e.g. the presence and distribution of words introducing further arguments, contrast, or conclusion) and generality/vagueness (e.g. words indicating information certainty or hedging/evasive language). The tool does not check grammar or spelling.

The three dimensions of analysis can be displayed in a single visualization providing a unique “fingerprint” of the text. For instance, it is possible to determine that a certain essay contains more connectives but fewer indicators of argumentative steps than other essays in the cohort under consideration. Starting from this global view the user can then zoom in on specific features and examine e.g., which connectives are used in the essay and how they are distributed across the text.

The tool does not provide an assessment of the quality of the essays. Its purpose is to help tutors and (writing) instructors to quantify their qualitative intuitions concerning certain characteristics of the text and to serve as a visual aid in the feedback process. The interpretation of the visualizations and of the specific measures is left to the user.

For questions and comments and in order to receive login credentials contact: rado@em.uni-frankfurt.de or I.Ivanova@em.uni-frankfurt.de.

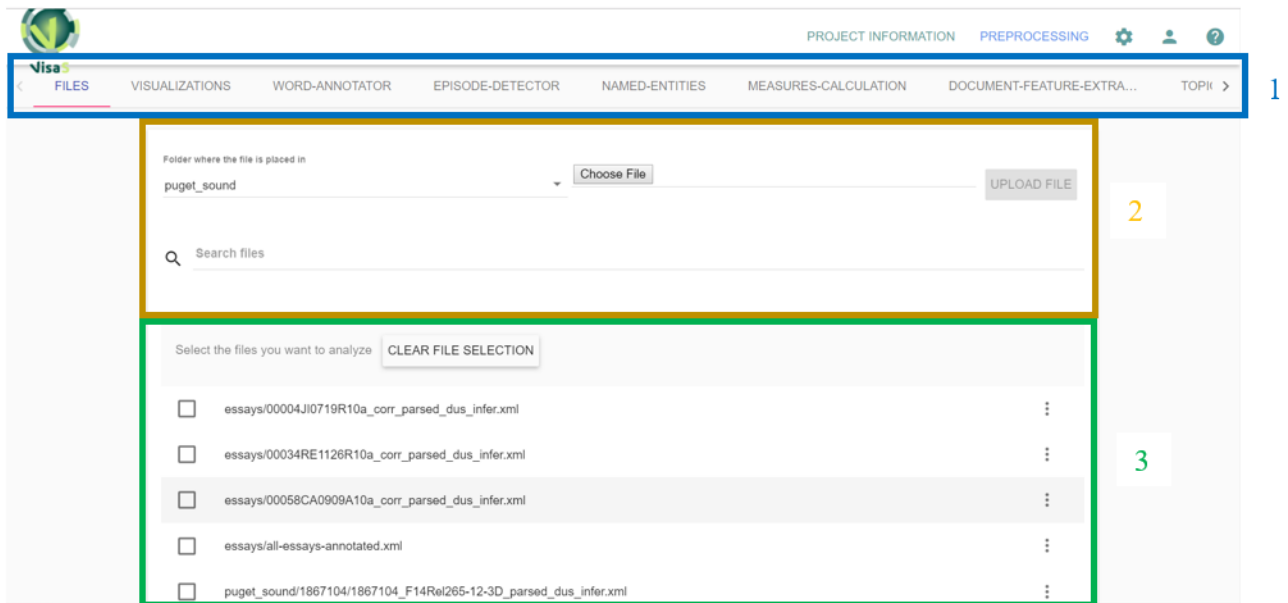
Table of Contents

Table of Contents	3
Getting started	4
Argumentation	7
Measures in the glyphs.....	9
I. Dimension Vagueness	11
I.1 Subdimension Information certainty.	12
I.2-3 Subdimensions Hedging and Non-hedging	12
I.4 Subdimension Event modality.....	12
I.5 Subdimension Generality	13
II. Dimension Coherence	13
II.1 Subdimension Topic development.....	14
II.2 Subdimension Sentence structure	16
II.3 Subdimension Reasoning	16
III. Dimension Sophistication	17
III.1 Subdimension Phrasing	17
III.2 Subdimension Emotion	18
III.3 Subdimension Vocabulary	19
Lexical Units.....	19
Left Panel.....	20
Central Panel	20
Right and Bottom Panels	21

Getting started

The VisaS tool is optimized for Google Chrome, using a different browser may result in limited functionality. To use the tool visit the link <http://visas.lingvis.io/> (the URL is subject to change for the external users) and enter your authentication credentials to log in. If you have trouble logging in then please let [us](#) know right away.

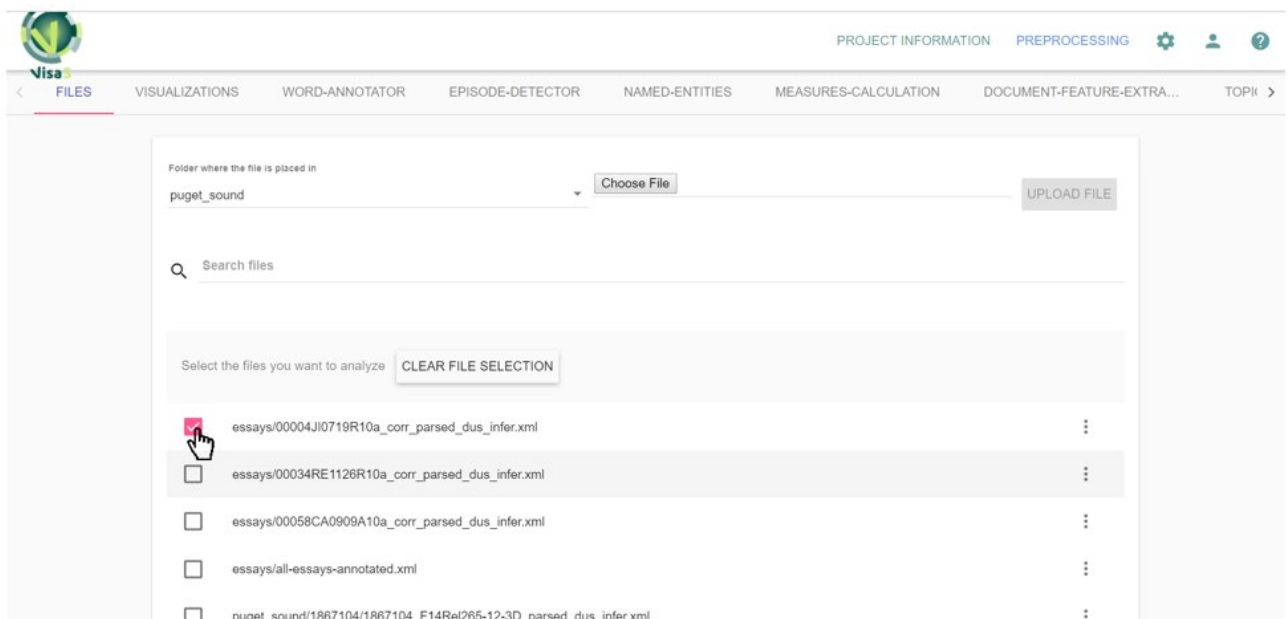
Once you are logged in, three fields load: the tool's navigation menu (1), the File Upload field (2), and the text corpus (3).

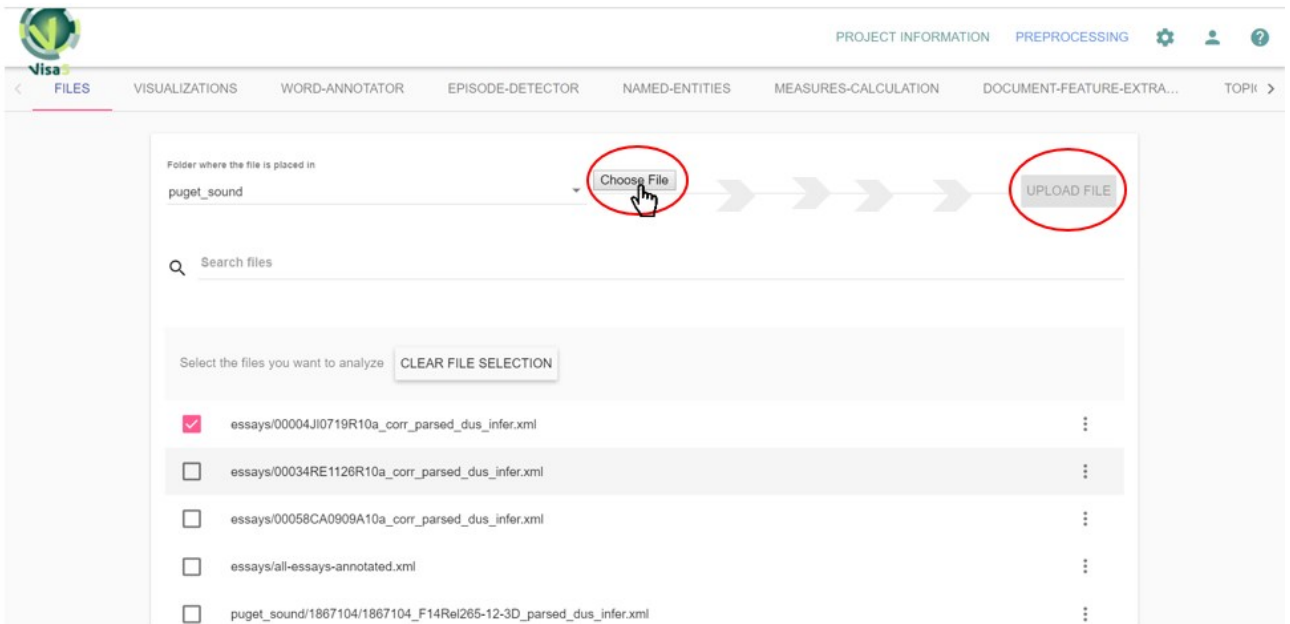


The current version of the tool contains a corpus of English essays in the fields of Linguistics, Literary Studies, etc. provided by students from Goethe University Frankfurt.

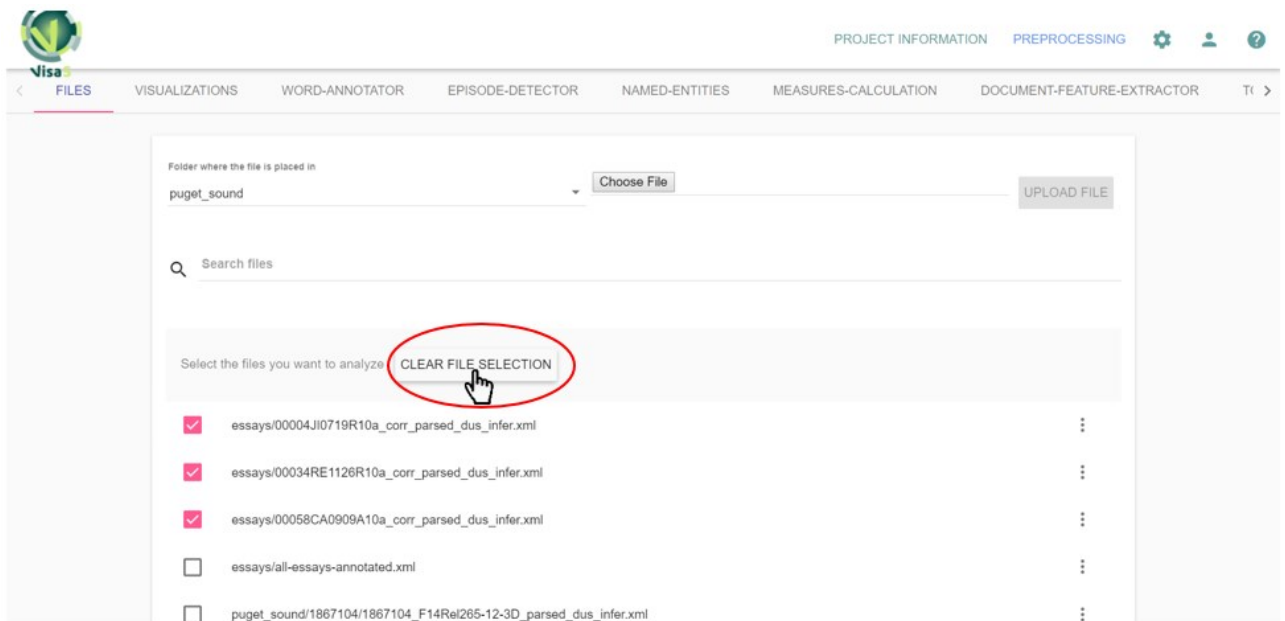
A text file may consist of a single text, or multiple texts by a single author, or texts written by different authors in a particular course/semester, etc., as indicated by the file names.

To select texts from the corpus, check the boxes in front of the file names.

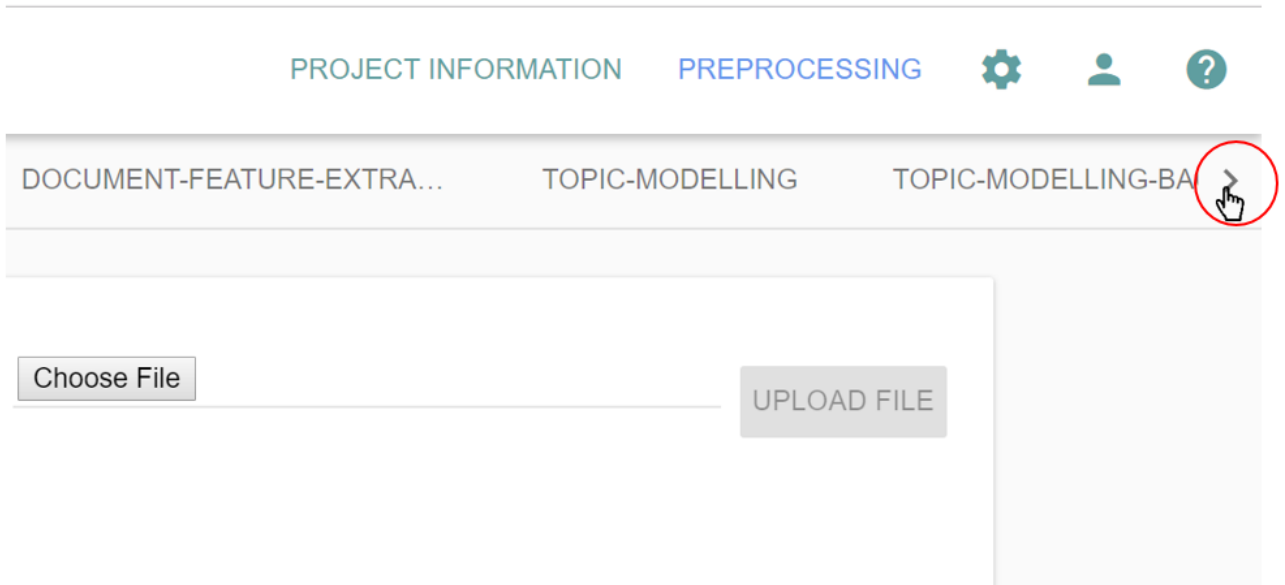




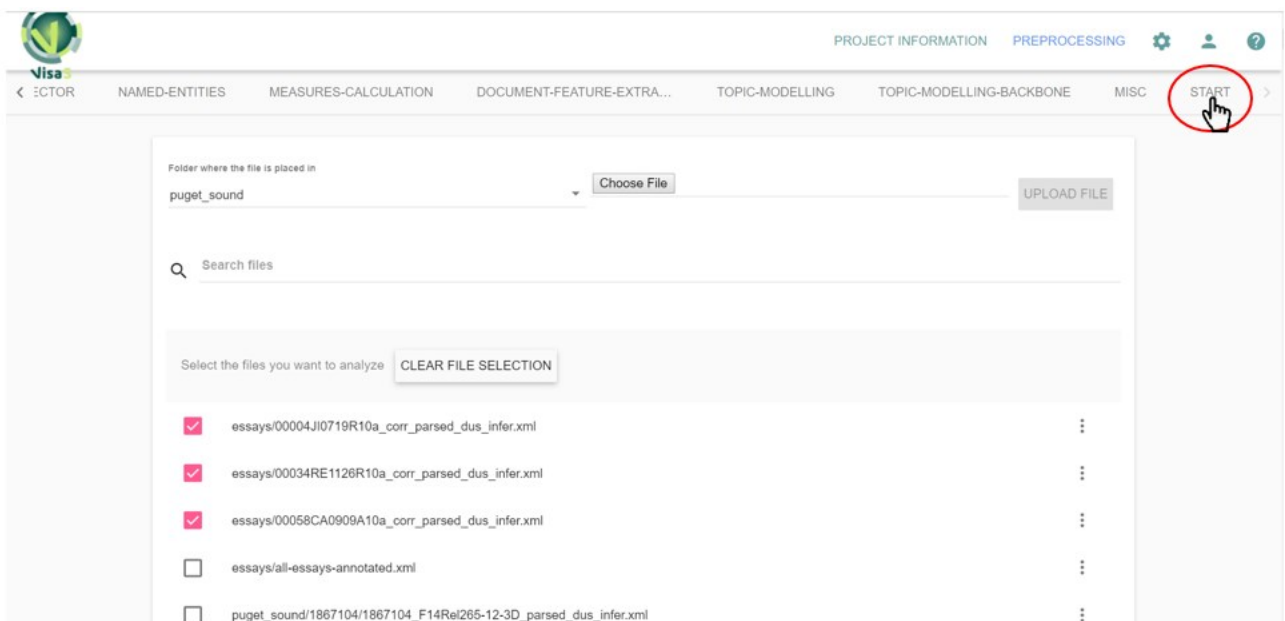
If you want to exclude a file from the analysis, uncheck the box in front of it. To uncheck all selected files, click on **Clear File Selection**.



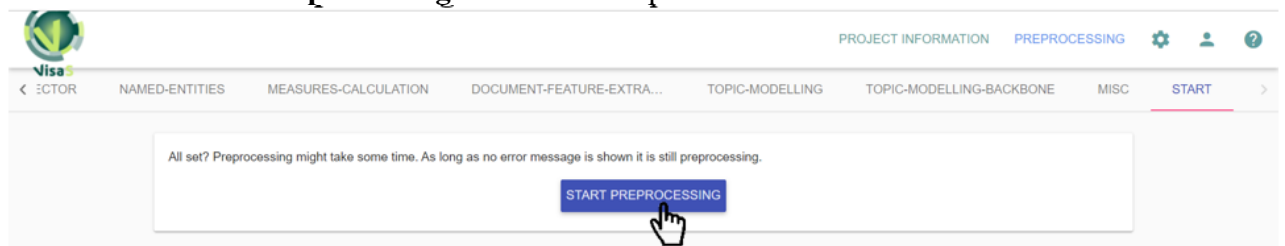
Once you have selected the text files you want to analyze, press the right arrow in the navigation menu to make your way to the Start button.



Then click on **Start** to launch the preprocessing of the selected files.

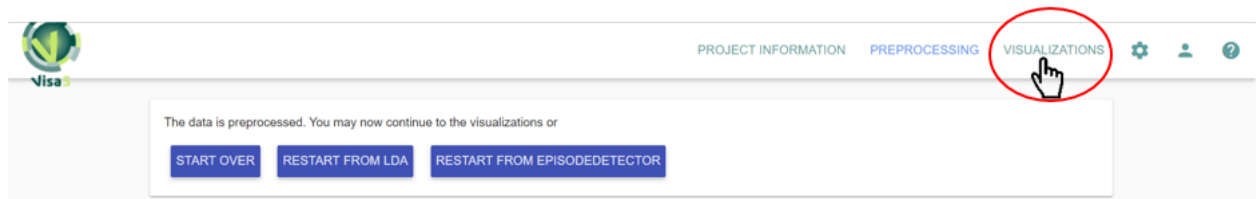


Then click on **Start Preprocessing** to confirm the process.

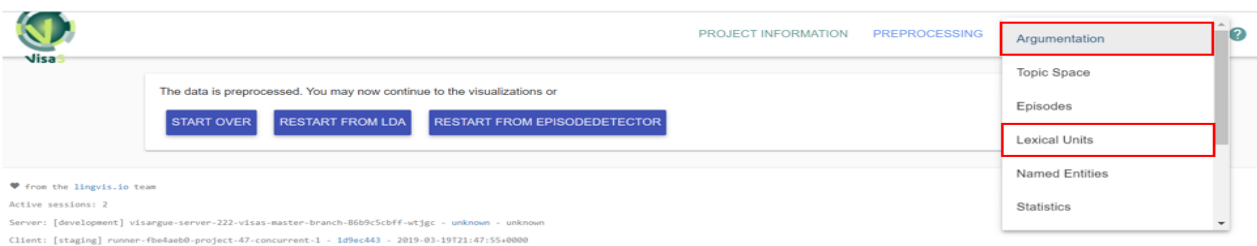


Once the data has been preprocessed you can either click on **Visualizations** from the navigation bar to view the types of visualizations available in the tool or go back to the file

selection by clicking on the **Start Over** button. When you click on **Visualizations**, a dropdown menu appears with the list of visualizations.



The visualizations you will be able to test in the current version of the tool are [Argumentation](#) and [Lexical Units](#).



The preprocessing uses the texts you have selected as units. In the [Argumentation](#) visualization units take the form of glyphs and in the [Lexical Units](#) visualization you can explore the content of the units and their individual linguistic characteristics.

Argumentation

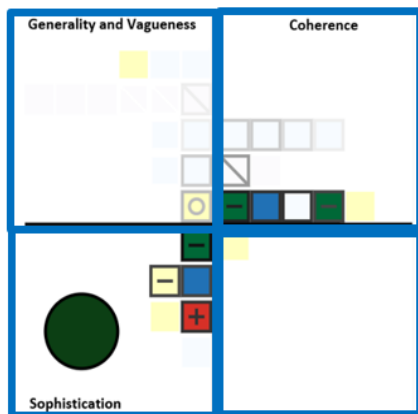
Essay-Glyph View

Visualization of text properties

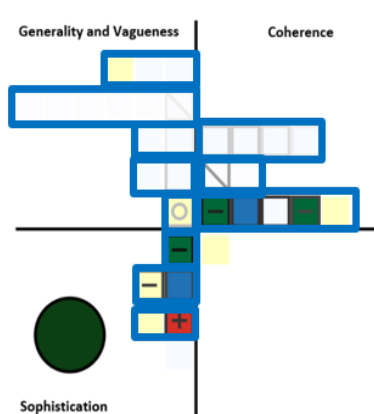
Glyphs are complex signs. Their components express a range of qualitative properties of a text or a number of texts. The values given represent scores compared to the other texts (authors, courses, etc.) included in the analysis. A glyph is thus always interpreted in comparison to other glyphs (other texts, other authors, etc.)

Glyphs consist of four quadrants corresponding to four dimensions in the data analysis. Three of these dimensions are currently used: Generality and Vagueness, Coherence, and Sophistication. Each dimension contains several subdimensions represented as rows; within a subdimension each square corresponds to one concrete measure. Subdimensions and measures are ordered by importance such that more important ones are closer to the x and y axes.

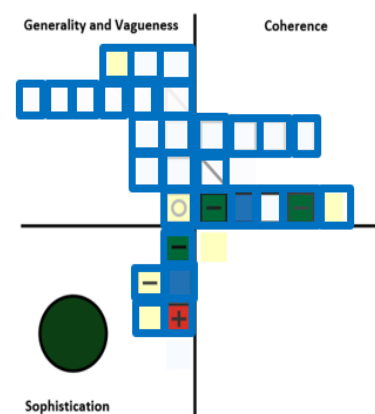
Dimensions



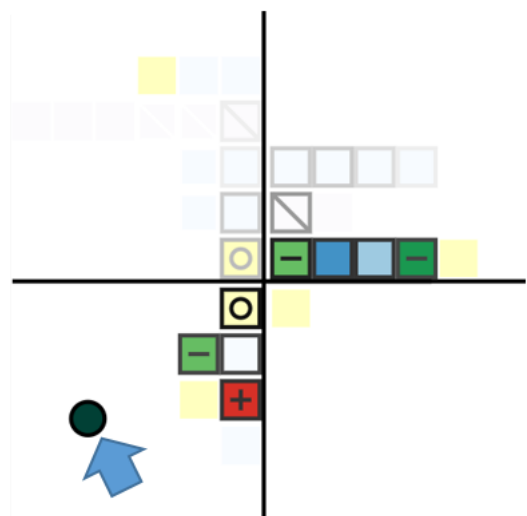
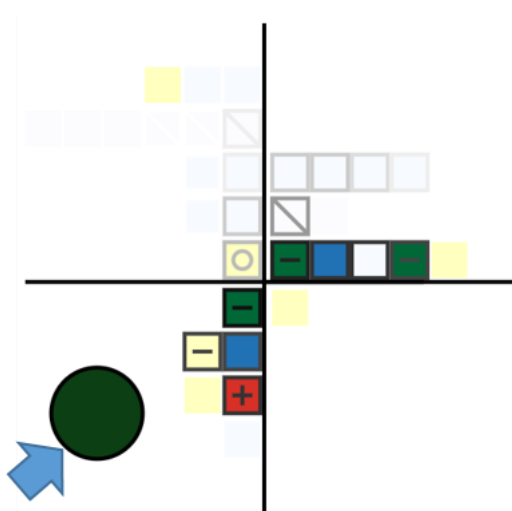
Subdimensions



Measures



The size of the circle on the bottom left of the glyph shows the relative amount of text being visualized.



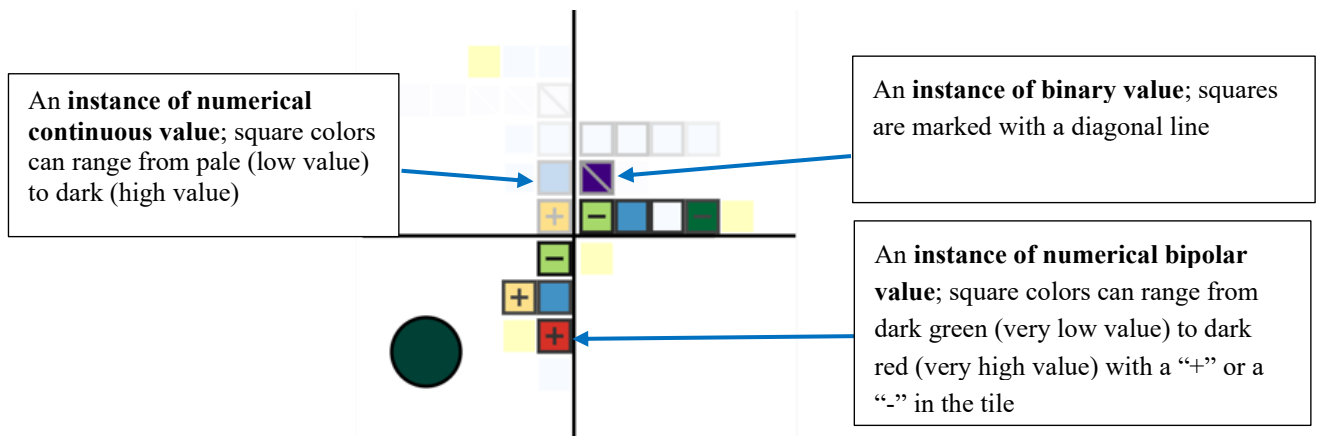
Three types of measures are used.¹

Numerical continuous values: The tool counts the relevant instances and calculates their average number per paragraph. Color scale: from pale (low value) to dark color (high value).

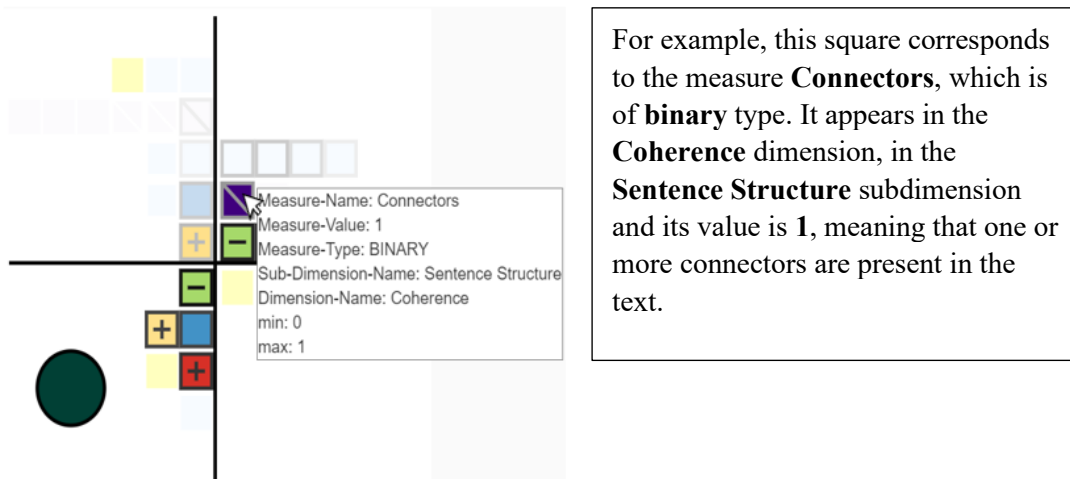
Numerical bipolar values are used for categories such as sentiment: negative sentiment is coded as -1, positive as 1, neutral as 0. The color scale is from dark green (very low values) to dark red (very high values). In addition, + or - indicates higher/lower values compared to the other glyphs.

Binary values are used if the choice is “yes” or “no”, indicating, e.g. if a connector is present or not, the zeros and ones are then averaged. In the visualization these squares are marked with a diagonal line.

¹ Learn more about how the measures are calculated in the ‘expert’ version of the manual.

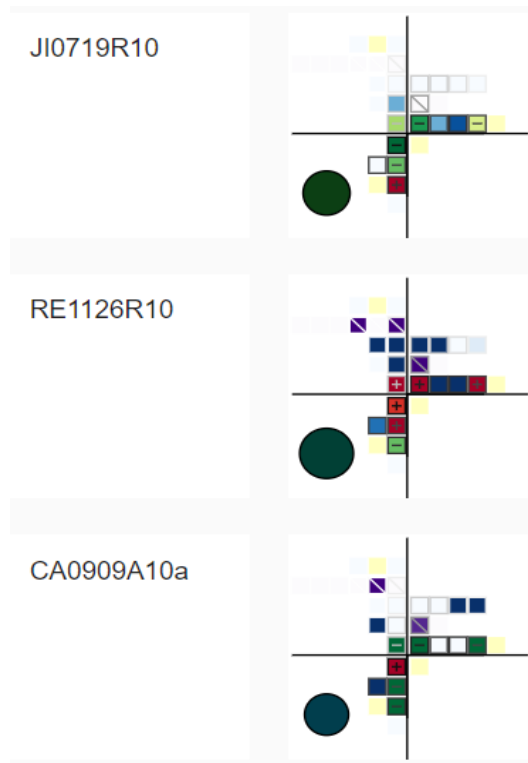


By hovering over the square you can see the dimension, subdimension, and measure represented by the square. The tool also gives you the type of measurement, the range of values across all glyphs, and the precise value for the glyph at hand.

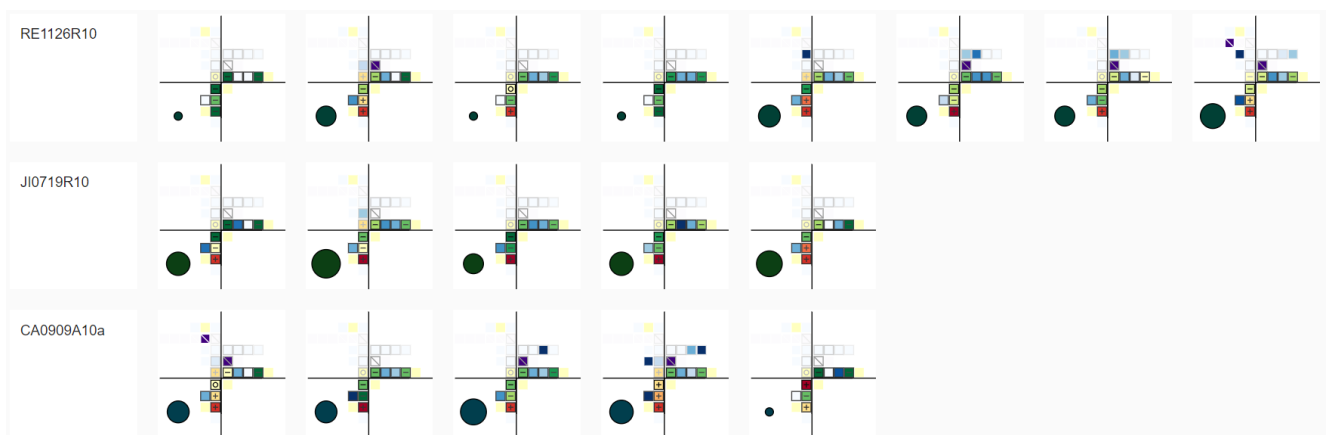


Measures in the glyphs

Assuming that you have loaded three texts for analysis, each by a different student, you will see three glyphs in the **Aggregated Speaker Glyphs** visualization. Each glyph corresponds to the whole text by that author.



The **Speaker/Utterance Glyph** table shows each student's text paragraph-by-paragraph, one row of glyphs corresponding to one author.

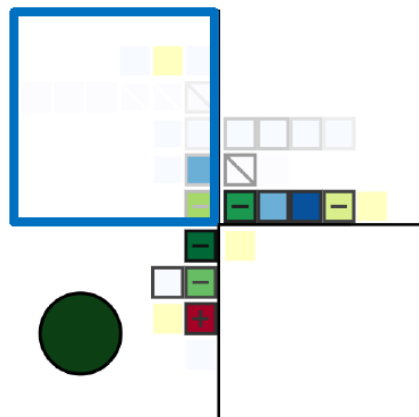


The **Speaker/Topic Glyph** table uses topics automatically identified by the tool. Keywords for the topics are provided across the top; a glyph represents one or more paragraphs on the topic by a given author.

	number limit manual arm	optical unable imagine full	signer means reflect agreement	resemble helpful precise lack	oversee fulfill similar explore	production express involve speed	fix set minimal refer	whisper shout place exception
RE1126R10								
J10719R10								
CA0909A10a								

The descriptions in the example below refer to a single glyph in the **Aggregated Speaker Glyph** visualization, i.e. a glyph capturing the complete text by an author.

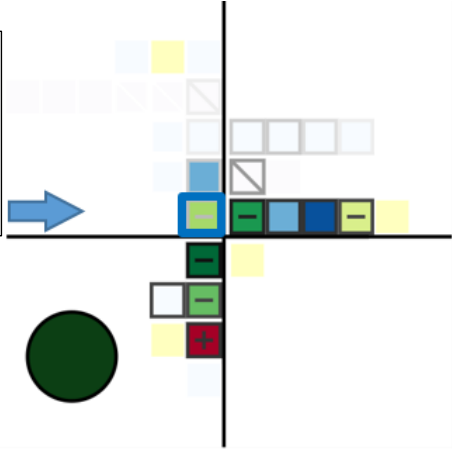
I. Dimension Vagueness (top left quadrant)



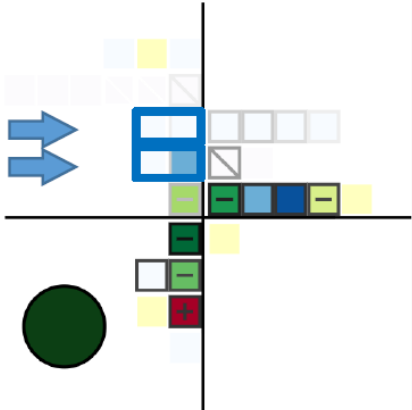
This dimension should help assess various aspects of clear and precise presentation.

I.1 Subdimension Information certainty (bottom row) is intended to capture how certain the author is about the information in the text. For that purpose, words expressing the epistemic values of *impossible*, *possible*, *probable* and *certain* are counted and weighted, then the result is divided by the total number of words in the text. To see which modals are used and where they occur in the text go to the visualization [Lexical Units](#) and click on *Modals* (bottom panel) and *Information certainty* (right panel).

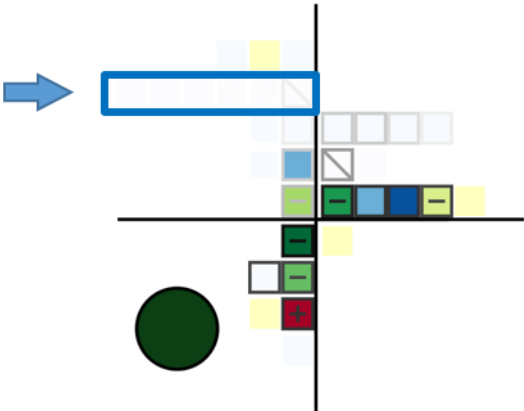
A “-” in the tile or a low value indicates low certainty compared to other glyphs in the visualization.



I.2-3 Subdimensions Hedging (second row) and Non-hedging (third row) should indicate whether the author provides information in a direct or indirect way.²



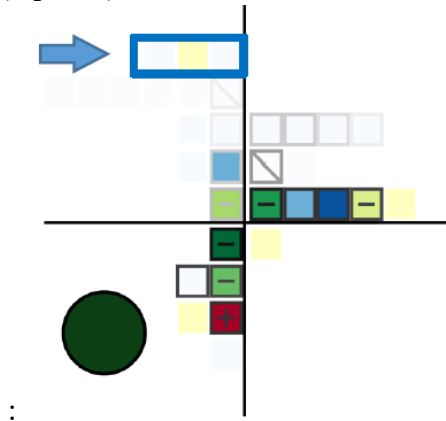
I.4 Subdimension Event modality (fourth row)



² They are based on a word list; these words are classified according to the epistemic categories in Information certainty but are weighted differently.

The categories in this subdimension are **permission, volition, obligation, alternative, reluctance, external constraint**. They are based on categories proposed in the linguistic literature and used in a software similar to VisaS. To see what indicators of event modality are used and where they occur in the text go to the visualization [Lexical Units](#) and click on *Event modality* (right panel)

I.5 Subdimension Generality (top row)

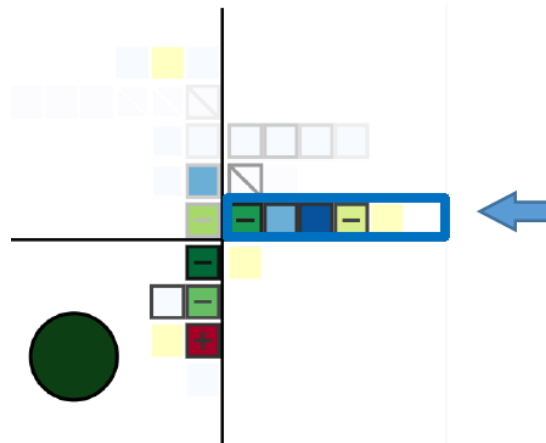


Generality comprises **Quantifiers, Reinforcement** and **General Verbs**, all based on word lists. This is a mixed batch: the use of quantifiers might indicate vagueness or evasiveness; general verbs are so-called light verbs such as *do* and *have* whose usage may be an indication of vagueness but also a matter of style. The **Reinforcement** category includes expressions indicating addition (*moreover*) but also emphasis (*indeed*).

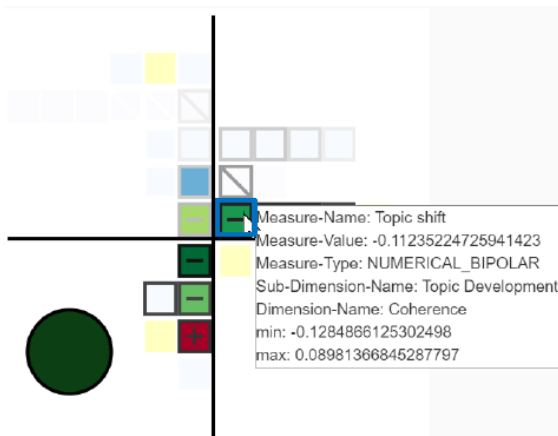
II. Dimension Coherence (top right quadrant)



II.1 Subdimension Topic development (bottom row)

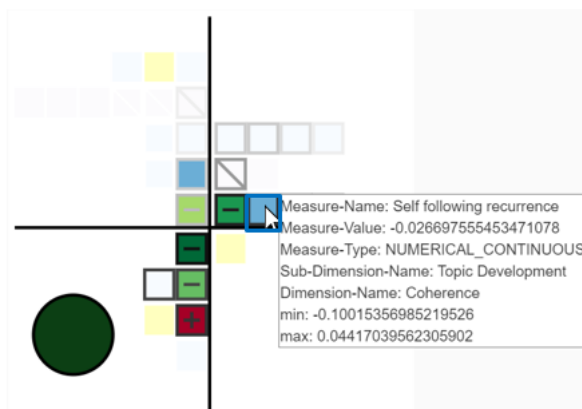


Topic shift is coded as 1 or -1 depending on whether a topic shift takes place or not, then an average value is computed for the text.



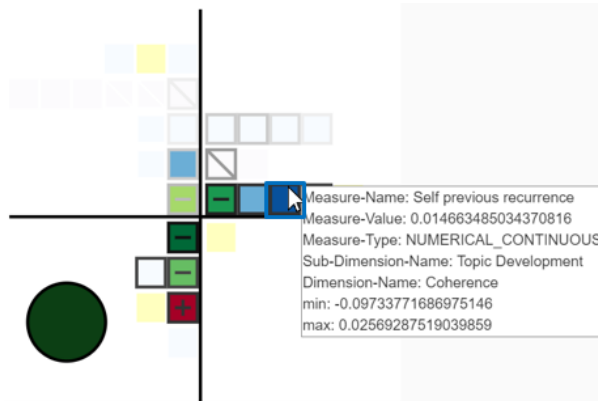
The green square with a “-” in the tile means that no topic shift is taking place in the text. The **min** and **max** values correspond respectively to the lowest and highest averaged topic shift values of the texts with which the current text is compared.

Self following recurrence measures how much of the content of the current paragraph is also contained in the following paragraphs of the same text, i.e. how much “influence” this paragraph has on the rest of the text.



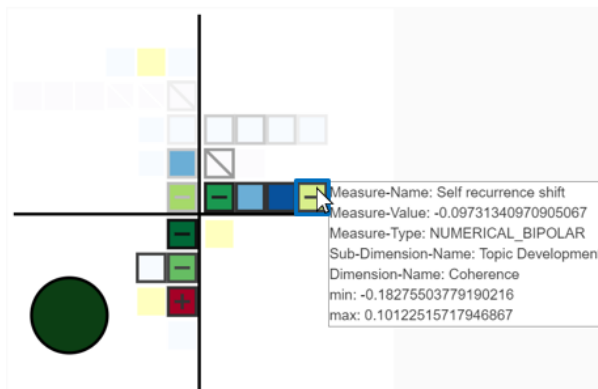
The negative value of the **self following recurrence** measure indicates that the content of a paragraph in the current text has very little “influence” on the rest of the text.

Self previous recurrence measures the relative amount of content recurrence of a given paragraph compared to all previous paragraphs in the same text, i.e. how much of the content of this paragraph is a repetition of what was said earlier in the text.



The positive value of the **self previous recurrence** measure indicates that the content of a paragraph in the current text repeats earlier parts.

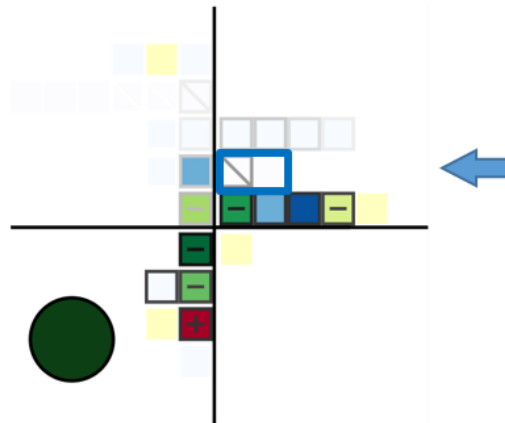
Self recurrence shift is a measure of the relation of the two previous categories: negative values indicate that a given paragraph is more relevant to the preceding part (e.g. as a summary), positive values show higher relevance to the following part of the text (e.g. as introducing a new (sub-)topic).



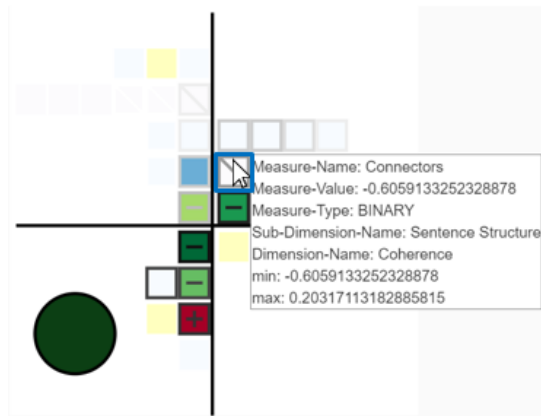
The negative value of the **self recurrence shift** measure indicates that the text represented in the glyph is more relevant to the preceding part, and/or suggests that the text might be a summary.

The last square stands for **Topic persistence**, it is currently not used.

II.2 Subdimension Sentence structure (second row)

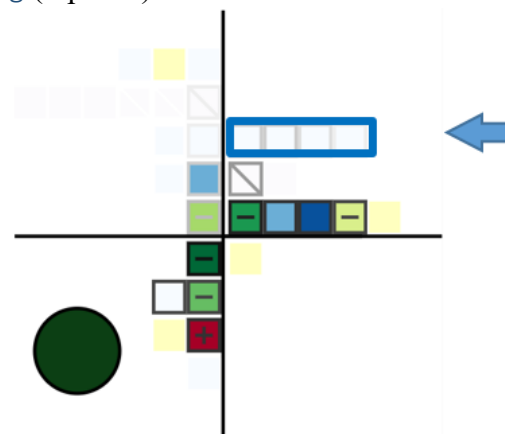


The measure **Connectors** receives the value 1 if the paragraph contains at least one of a list of connectors, otherwise it is 0. To see which connectors are used and where they occur in the text go to the visualization [Lexical Units](#) and click on *Connectors* (bottom panel).



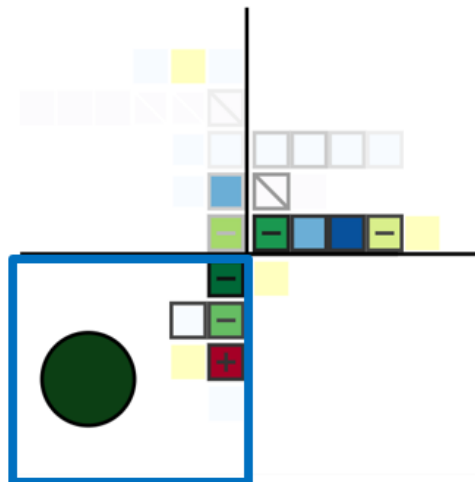
The measure **Arguing** is based on a speech act annotation that is currently not used in VisaS.

II.3 Subdimension Reasoning (top row)



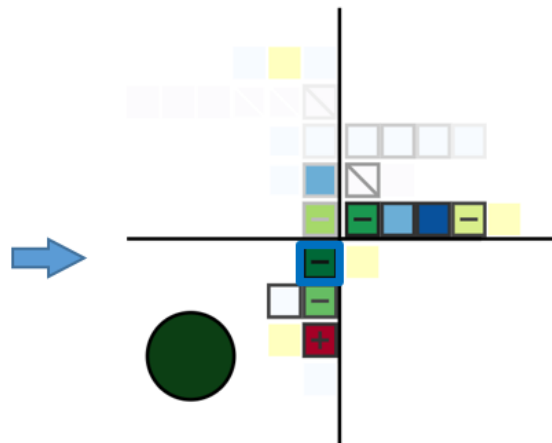
This subdimension captures elements of the argumentation structure of the text. The measures used are **Reason**, **Conclusion**, **Condition** and **Consequence**. They are all calculated on the basis of word lists; each sentence in a paragraph receives the value 0 or 1 for each of these measures based on the absence/presence of the relevant connector type, then an average is calculated for the paragraph. To see which argumentation-related connectors are used and where they occur in the text go to the visualization [Lexical Units](#) and click on *Causal argumentation* (side panel).

III. Dimension Sophistication (bottom left quadrant)



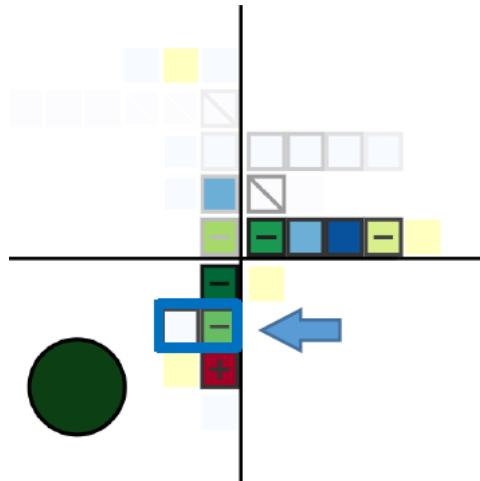
This dimension should capture stylistic properties of the text.

III.1 Subdimension Phrasing (top row)



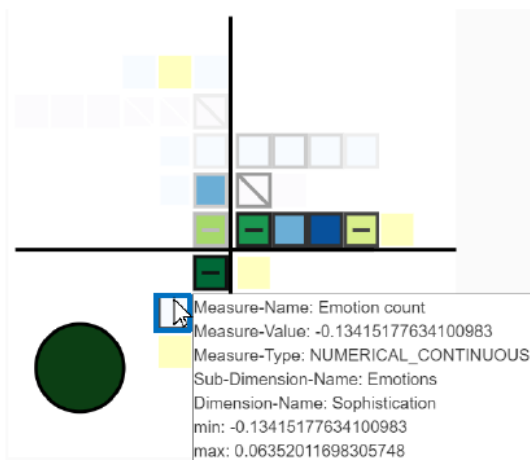
Average sentence complexity is calculated by assigning the value 1 to complex sentences (those including a subordinate clause) and -1 to simple sentences, then averaging. Despite being in the category *Sophistication* there is no simple interpretation of this measure correlating higher values with more sophisticated writing; it merely reflects the proportion of complex sentences in the text. An extremely low value might indicate simplistic writing but may also be the sign of a clear, simple style.

III.2 Subdimension Emotion (second row)



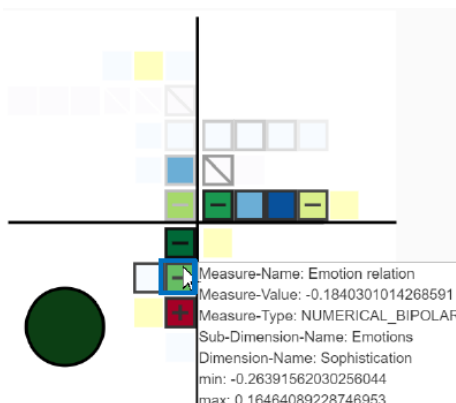
Emotion indicators in the text are marked by words such as “major,”/“minor,,
“,”“lack,,/“true,,/“advantage,, etc.

Emotion count is the total number of emotion words divided by the number of words. To see which expressions were counted and how they were evaluated for this measure to the visualization [Lexical Units](#) and click on *Emotion* (side panel)



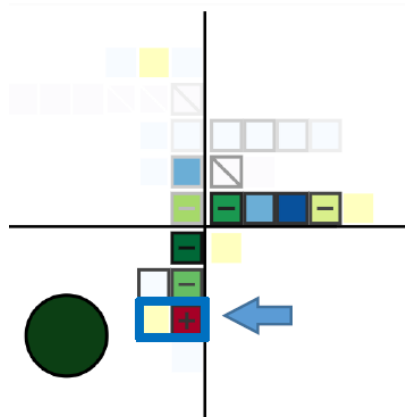
The negative value of the **Emotion count** measure indicates that only few emotion words appear in the current text.

Emotion relation is calculated by subtracting the number of negative emotion words from the positive ones, thus it has a positive value if more of the half of emotion words in the text is positive.

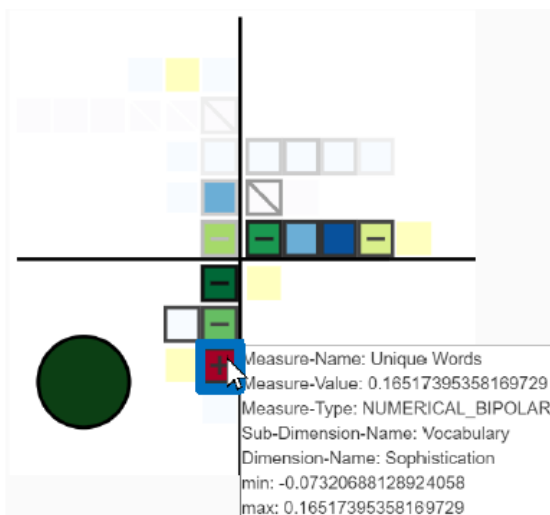


In this case, the negative value of the **Emotion relation** measure indicates that more than half of the emotion words are negative.

III.3 Subdimension Vocabulary (bottom row)



##Unique words measures lexical diversity: it assigns the value 1 to sentences that contain new content words compared to the other sentences. It should be a measure of words in this text that do not occur in other texts.



The positive value of the **Unique Words** measure indicates that the author of the current text has used unique words.

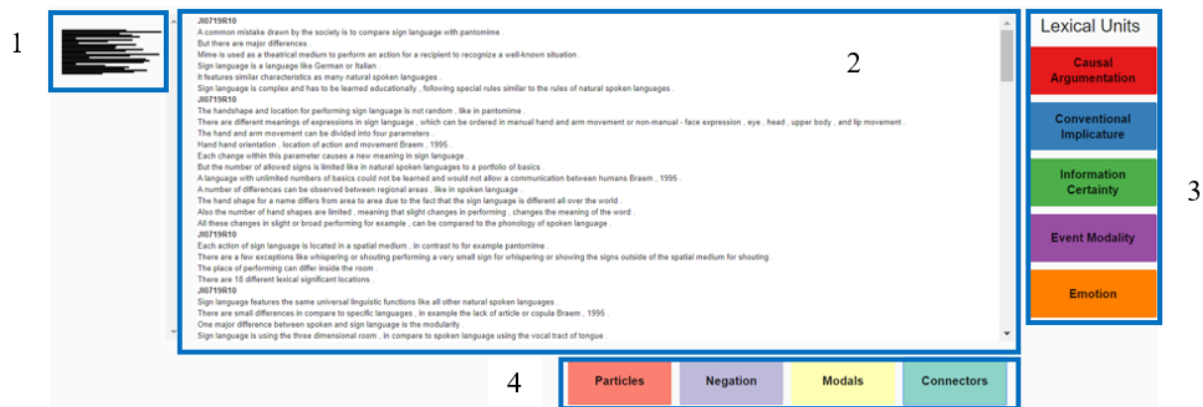
Single mentions is currently not used.

Lexical Units

Lexical-Unit Annotation Explorer Basic Data Exploration View

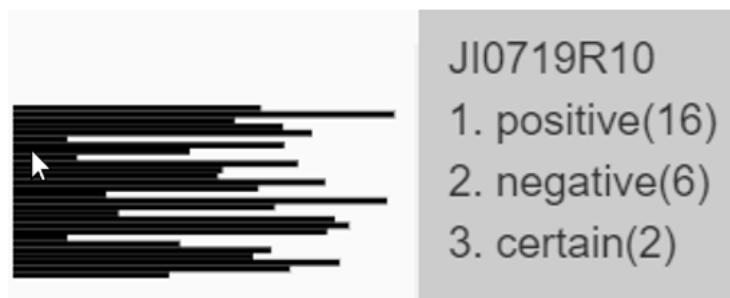
This view presents an interactive component for the exploration of the linguistic annotations in the text. Using interactive selection and highlighting, you can analyze pre-defined and automatically annotated features. You can track, for instance, which connectors are used in a given essay and where they occur, which is useful for assessing the argumentation steps used in the essay. The annotations are subdivided into semantic categories.

The view consists of 4 panels: left (1), central (2), right (3), and bottom (4).



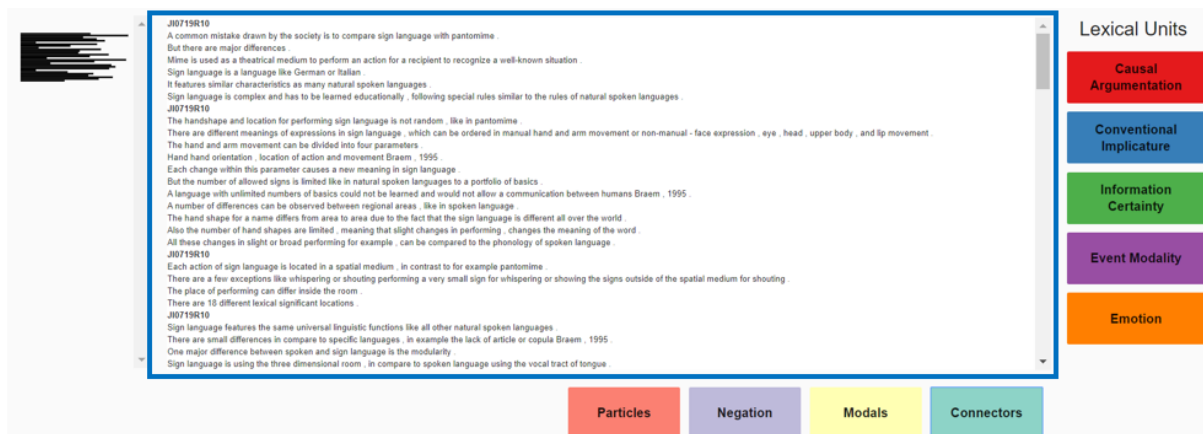
Left Panel (marked (1) in the figure above)

The left panel provides a bird's eye view of the texts you are currently analyzing and offers an interactive navigation. When you hover over the lines, a tooltip appears on the screen. This tooltip contains a label for identifying the text (e.g. *JI0719R10*), a list of the linguistic features present in the text (e.g. *positive, negative, certain*), and the number of times each of these features appear in the text (e.g. 16 indicators of positive emotion, 6 indicators of negative emotion, 2 indicators of certainty).



For example, the following excerpt contains a sentence with a single indicator of a positive emotion (highlighted). “Spoken language is more common than sign language. One of the aspects could be, **that it is easier to understand and to communicate.**”

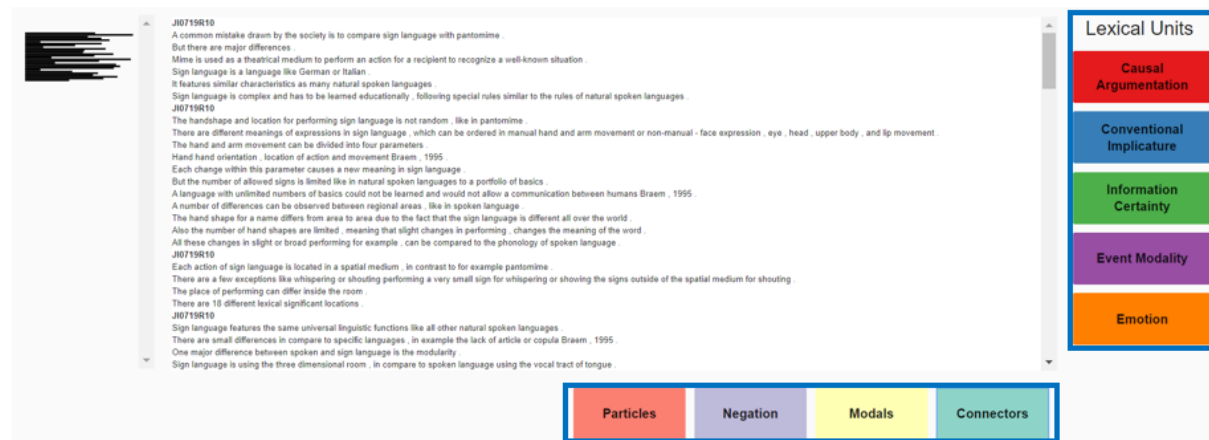
Central Panel



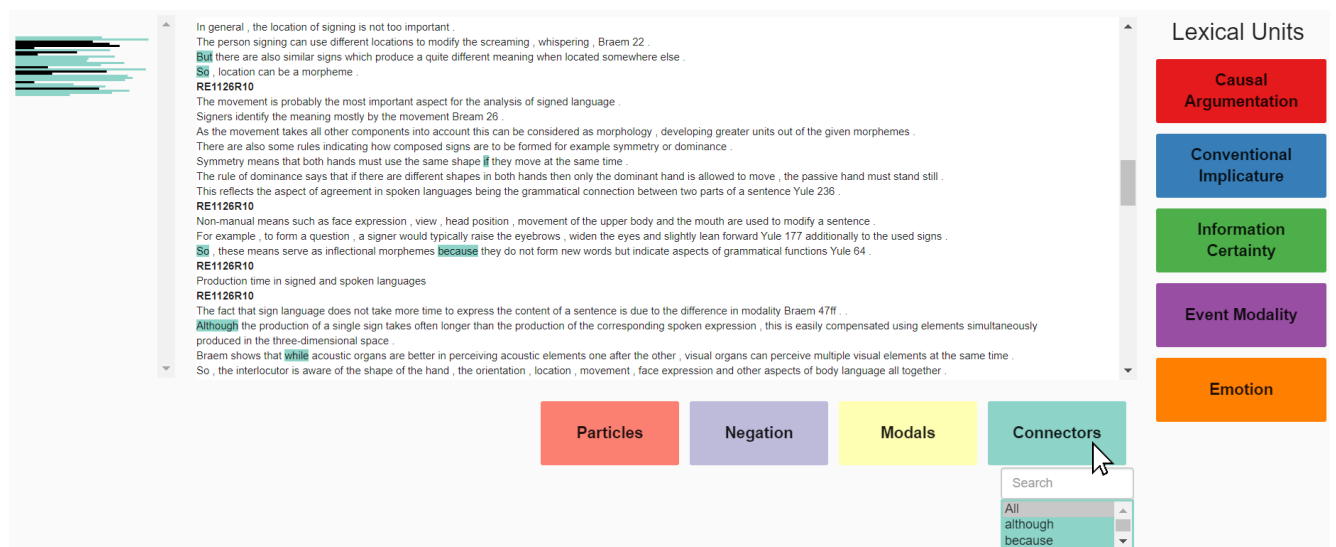
The central panel displays all analyzed texts one after the other. Each sentence appears on a new line, paragraph breaks are indicated by the label for identifying the text. When multiple texts are displayed, you can either use the scrollbar to look through all the texts or simply

click on the lines in the left panel, which will redirect you to the selected text in the central panel.

Right and Bottom Panels



The right and the bottom panels contain a list of clickable semantic categories: Causal Argumentation, Conventional Implicature (*currently not used*), [Information Certainty](#), [Event Modality](#), [Emotion](#), Particles (*currently not used*), Negation, [Modals](#), and [Connectors](#). When you select a category, a dropdown box appears. The box contains a list of subcategories or words which are indicators of the semantic category. These subcategories/words can be selected either individually or simultaneously. To see what indicators are used for each category and where in the text they are highlighted, click on the categories.



For example, clicking on **Connectors** opens a dropdown box with a list of connectors. **Connectors** are conjunctions such as *therefore*, *whereas*, *because*, *while*, *although*, *thus*, etc. that indicate the type of relation (e.g. *contrastive*, *causal*) that holds between different threads of meaning in the sentence or between the sentences. With this semantic category, you can assess how well the author has presented their argumentation. You can either select one or more connectors to see if they are present in the text or can select **All** to see all the connectors which are present and where in the text they appear. The places in which the connectors appear are also highlighted in the left panel, which can be used to navigate through the texts.