

TransVis: Integrated Distant and Close Reading of Othello Translations

Mohammad Alharbi, Robert S Laramée, and Tom Cheesman

Abstract—Studying variation among time-evolved translations is a valuable research area for cultural heritage. Understanding how and why translations vary reveals cultural, ideological, and even political influences on literature as well as author relations. In this paper, we introduce a novel integrated visual application to support distant and close reading of a collection of Othello translations. We present a new interactive application that provides an alignment overview of all the translations and their correspondences in parallel with smooth zooming and panning capability to integrate distant and close reading within the same view. We provide a range of filtering and selection options to customize the alignment overview as well as focus on specific subsets. Selection and filtering are responsive to expert user preferences and update the analytical text metrics interactively. Also, we introduce a customized view for close reading which preserves the history of selections and the alignment overview state and enables backtracing and re-examining them. Finally, we present a new Term-Level Comparisons view (TLC) to compare and convey relative term weighting in the context of an alignment. Our visual design is guided by, used and evaluated by a domain expert specialist in German translations of Shakespeare.

Index Terms—Text visualization, Othello, Parallel Translations.

1 INTRODUCTION AND MOTIVATION

TEXT visualization is a popular subfield of information visualization due to the rapid increase in digital text data over the last two decades [3], [53]. In this project, researchers with an expert background in the Arts and Humanities prepared 38 translations of Shakespeare’s play *The Tragedy of Othello, the Moor of Venice* (1604). The translations were originally written over a span of 244 years from the Christoph Martin Wieland translation [88] in 1766 to the Christian Leonard translation [54] in 2010. Our data set contains the 38 translations as well as the English base text of a sample of the full play –Act 1 Scene 3.

Based on the notion of close and distant reading of texts [11], [58], we attempt to create a novel interactive visual design that combines distant and close views of the parallel translations of Shakespeare’s *Othello*. Guided by the visual information seeking mantra [78], inspiration from previous work on this topic, and close collaboration with the domain expert we derive six requirements and five tasks that our application must support (discussed in Section 4). The proposed visual design is guided closely and reviewed by a domain expert from Arts and Humanities.

Contributions: In this paper, we contribute the following:

- We support integrated distant and close reading in the same view and implement them with smooth zooming and panning.

-
- *Mohammad Alharbi is with the Department of Computer Science at Swansea University, United Kingdom.
E-mail: m.alharbi.508205@swansea.ac.uk.*
 - *Robert Laramée is with the School of Computer Science at the University of Nottingham, United Kingdom.
E-mail: robert.laramée@nottingham.ac.uk.*
 - *Tom Cheesman is with College of Arts and Humanities, Swansea University, United Kingdom.
E-mail: t.cheesman@swansea.ac.uk*

- A novel visual design that supports comparison of an arbitrary number of parallel translations.
- Customized mechanisms for rapid and interactive filtering and selection of a large number of German translations of Shakespeare.
- Interactive and dynamic analysis of similarity metrics to support comparisons and analysis of customized parallel translations.
- Examples, detailed observations, a case study, and domain expert feedback from a specialist in German translations of Shakespeare.

The rest of this paper is organized as follows: Section 2 introduces and defines the most important terms used in this paper and introduces the parallel data as well as the similarity metrics. Section 3 discusses previous work related to our approach and the challenge domain. Section 4 outlines the design requirements and tasks that our approach supports. Section 5 introduces our proposed application components. Section 6 provides the domain expert feedback and case studies. In Section 7, we discuss the study outcome and report on the design process utilizing guidance provided by Sedlmair et al [77]. We finish with conclusions and future work directions.

2 DEFINITIONS AND BACKGROUND

This section provides background information on terminology, the text data, and translation meta-data.

- **Definitions:** In this section, we explain the notions of close and distant reading. **Close reading** generally defines the process of carefully reading word-for-word and interpreting a passage to develop a deep understanding of the ideas contained in the text [11]. Close reading signifies the critical analysis of small and specific components of the text over the general theme. Close reading may involve annotation and highlighting techniques to increase the comprehension

process. In a literary context, close reading is defined by Nancy Boyles [11] as “*reading to uncover layers of meaning that lead to deep comprehension.*” The practice of close reading is inherently subject to the reader, and interpretations of the text vary according to readers and context [56].

On the other hand, **distant reading** aims to provide an overview of the text by moving from an in-depth exploration of the individual components of the text to presenting the global features of the text(s) [46]. In contrast to close reading, distant reading is technically a more objective process because most of the context is hidden and the reader is left with the result of computationally and analytically abstracted visualization.

Throughout the paper, some special terms are used. A **segment** is a meta-data object. A segment can be any continuous sequence of words within a text, but generally, it is a meaningful unit. The text of Othello is a play for theatrical performance. Like most play texts, it contains three kinds of segment: ‘speeches’ (words to be spoken by actors), ‘speaker identifiers’ (words indicating the character in the play who speaks), and ‘stage directions’ (words which instruct the director and actors). A speech is always preceded by a speaker identifier. A speech can consist of one to many sentences, and one to many words. In our dataset, all three kinds of text are predefined as segments.

An **alignment** is a meta-data object which links a given translation segment with its corresponding text in the English base text. In our dataset, alignments have been created (in a machine-assisted manual process) between each segment in the English base text, and each segment in all the translations which has a meaning which corresponds to the base text segment. Some translations omit, transpose, and add new words, sentences, and even speeches, so aligning is a complex task. However, the majority of translations in our dataset are ‘faithful’, ‘close’ and complete translations. With those, making speech-by-speech alignments is relatively straightforward.

• **Description of Parallel Translation Data:** A group from the Arts and Humanities working on a project called “Translation Arrays: Version Variation Visualization” [20] collected 55 translations of Shakespeare’s play Othello (1604) Act 1, Scene 3 into German. The translations span 1766–2010. So far, 38 translations of the collection were optically scanned from paper prints, corrected from OCR errors, segmented, and aligned with the English base text to create a parallel corpus. The set of translations we study was collected over a time-span of 2-3 years from various sources, such as libraries, second-hand book-sellers, archives, theater publishers and theater companies. The translation data is stored in XML format on the project’s website [19].

Each `<document>` node in the dataset XML file represents a German translation of the base English text and is associated with a variety of meta-data such as, authors `<authortranslator>` and description of the translation `<description>`. The `<document>` node consists of three nodes: `<doccontent>`, `<segmentdefinitions>` and `<alignments>`. The base English text is also represented as a node of `<document>`, and does not have the `<alignments>` sub-node.

1. `<doccontent>`: contains the actual content of the

document using a number of `<blockquote>` nodes. Each `<blockquote>` has a number of `<q>` nodes which contain the actual text segments.

2. `<segmentdefinitions>`: stores the meta-data associated with each segment such as segment ID, length, speaker, etc.

3. `<alignments>`: is included in all documents except the base text. It consists of a number of `<alignment>` nodes which match a given segment of a translation with the corresponding segment of the base text using the *segment ID* elements stored in the `<segmentdefinitions>` node.

• **Similarity Metrics “Eddy and Viv”:** The Eddy and Viv metrics were introduced by Cheesman *et al.* [21] to quantify how a given base segment (English in our case) is interpreted and translated between parallel texts (German in our case). Eddy characterizes the translated segments in terms of distinctiveness. A higher Eddy value indicates higher dissimilarity from other translations.

The word forms are important at this stage. In the previous work, the formulations of Eddy and Viv do not consider advanced linguistics algorithms to reduce inflectional forms of words such as lemmatization or stemming [52]. This is due to the nature of the German language which is considered inflected. Special challenges appear with German Shakespeare texts due to the use of antiquated language and poetic orthography. However, we build up a lemmatization dictionary for our corpus using Cascaded Analysis Broker “CAB” [45] which is developed for the German Text Archive (Deutsches Text Archiv, DTA) [10]. CAB is an HTTP-based web service morphological normalization tool developed for historical German text especially for the 18th and 19th centuries.

As a dimensionality reduction algorithm, each segment is represented by a fixed length of vector of word weights TF-IDF (term frequency–inverse document frequency) [66], [71]. Then, the similarity coefficient between segments vectors can be obtained by the Euclidean Distance between each pair of segments. Euclidean distance is usually the default metric used to measure the distance between two points or vectors. It is the default distance metric used with the K-means algorithm [39].

After obtaining the similarity values between each pair of aligned segments, a weight value “Eddy value” is computed by averaging the sum of similarity values between each pair, such that:

$$Eddy(S_i^j) = \frac{\sum_{k=1}^n \|S_i^j - S_i^k\|}{n} \quad (1)$$

where S_i^j denotes segment i in translation j and n denotes the number of translations.

Viv, on the other hand, is the average pairwise distance between every two segments projected on the base segment, also known as the diameter of a cluster [70]. Viv represents the stability of the base segment. A high Viv value indicates low stability (high variability) which means the segment’s translations vary considerably between authors. We compute the Viv value for a given base segment i as:

$$Viv(i) = \frac{\sum_{k=1}^n Eddy(S_k^i)}{n} \quad (2)$$

High Eddy and Viv values are interesting to arts and humanities researchers because if a translation has high Eddy values, it indicates that the translator is working in a more unusual, possibly a more creative way relative to others, maybe interpreting the text in a new way. This might be due to circumstances such as historical changes in the language and/or the culture, as a result of political, economic and social change for example. It may be due to an individual translator developing their own new approach to the translation task. Or it may have to do with a new market developing for a new kind of text –in this case, new kind of theatrical drama. High Viv values indicate which base text segments are associated with variation among translations, which enables research into the textual factors (such as complexity, ambiguity, polysemy, semantic salience or affective intensity) which may provoke translators to deviate from one another.

Eddy and Viv metrics can be computed interactively and dynamically, that means every user customization derives new similarity metrics.

3 RELATED WORK

McNabb and Laramée [55] survey the surveys in information visualization literature and feature a text-focus category in their classification. Alharbi and Laramée [3] also review the existing surveys that specifically review text visualization literature. They include and classify 13 surveys that review text visual analytic approaches. Wanner *et al.* [85] review the literature of text event detection techniques that are used along with visual analytics approaches. Also, Kucher and Kerren [53] survey the literature of the text visualization and classify them on a novel text visualization taxonomy. Jänicke *et al.* [46], [47] are the only surveys that consider Digital Humanities as well as the data visualization community.

We discuss general comparison as a task and the design that supports it based on Gleicher *et al.* [29] in Section 3.1. Then, we group the related literature into four groups. The first category in Section 3.2 presents the general design that supports comparison of text in general. The second group in Section 3.3 discusses the designs that support visual comparison between parallel text. The third group in Section 3.4 presents visualization solutions that enable digital humanities tasks. The fourth group in Section 3.5 introduces previous visualizations of Shakespeare’s Othello.

In Table 1, we summarize the visual comparative approaches that facilitate parallel text comparison tasks. We indicate what type of documents they study, the maximum number of documents viewed in parallel in both close and distant views, the visual designs used in the distant views, as well as the language studied. If the meta-data is not provided explicitly, we extract it from the examples provided.

3.1 Visual Comparison as a General Task

Comparison is included in most task taxonomies [13], [68], [90]. It facilitates the exploration of the data in order to understand the similarities or differences between comparison elements [4], [28].

There are many approaches developed to perform comparative tasks. However, users and systems can often perform comparative tasks even if the systems are not developed in the context of comparison [28]. Gleicher *et al.* [29] group comparative visualizations into three representations: juxtaposition which shows objects side-by-side, superposition which overlays objects in the same visual space, and explicit representation of relationships. We can distinguish between the three categories in different ways. The separated object’s design relies on the memory of the user to conduct a complete comparison and there is no correspondence between them. Overlay designs use the same visual coordinate system to layout objects and proximity is needed to represent the connection. Explicit encoding facilitates computational tasks to investigate relationships. Gleicher *et al.* believe that interaction techniques such as brushing and linking can be helpful when applied to facilitate visual comparison. Also, using animation to show, for example, transitions can be helpful in understanding the connection between objects. Animation approaches can be useful to aid comparison between related objects, however, they can be problematic when not implemented carefully [36].

3.2 Visual Designs for Visual Comparison of General Text

There are approaches designed to support the analysis of a single document which, however, can be extended to facilitate comparison between multiple documents. Fingerprinting approaches [50], [61] are used to highlight semantic text properties on different hierarchy levels such as the development of relationships between characters in literature. The Text Variation Explorer [79] provides linguistically-assisted visualization to examine a user-selected text window. Vari-focalReader [51] presents an interactive multi-layer visual design based on the hierarchy level of a text, such as chapters, pages, and sentences. It incorporates multiple visual presentations to support the analyst exploring the document, such as bar charts, pictograms, and word clouds.

Many visualization systems are designed to analyze collections of documents without explicit support of inner relationships between documents. For example, Brehmer *et al.* [12] present an open-source platform that analyzes user-uploaded documents. It also allows the user to create custom visual layouts. Also, there are many techniques that visualize results of queries, such as Sparkler [35] which plot the results on a radar-like view and the search-engine similarity (SES) tool [81] which visualizes the results of multiple web search engines using multiple views.

Our work is different from the aforementioned work. We present a specially customized visualization of explicit parallel texts that are strongly related to each other, i.e. versions or editions of the same text. In the next section, we examine the related work in the area of parallel text.

3.3 Visual Designs for Comparison of Parallel Texts

There are a number of different visual representations that can facilitate visual comparison between multiple texts. The most common layout to visualize parallel texts is juxtaposition. For example ItLv [57] combines a timeline chart with multiple bar charts stacked vertically to represent documents. Another example is the Versioning Machine tool

References	Source of text studied	Max number of documents viewable simultaneously		Visual design of distant reading view	Languages
		Close Reading	Distant Reading		
Ribler and Abrams [63]	Programming codes	-	Arbitrary	Patterngram	English
Monroy <i>et al.</i> [57]	Literature (Don Quixote)	1	5	Multiple bar charts	Spanish
Schreibman <i>et al.</i> [74]	Poetry	Arbitrary	-	-	English
Jong <i>et al.</i> [33]	Unspecified documents	9	14	Multiple views+pixel-based visualization	English
Collins <i>et al.</i> [22]	U.S. Circuit Court Decisions	1	13	Parallel coordinates+tag cloud	English
Büchler <i>et al.</i> [16]	Ancient Greek text	3	-	-	Greek
Welsh and Hooper [84]	Newton Alchemical corpus	2	-	-	English
Geng <i>et al.</i> [27]	Literature (Othello)	8	8	Parallel coordinates	German
Behrisch <i>et al.</i> [8]	News	2	33	Heatmap matrix	English
Howell <i>et al.</i> [38]	Literature (The Secret Scripture)	2	-	-	English
Jänicke <i>et al.</i> [48]	The Bible translations	2, 7	2	Text Re-use grid, Dot Plot view	English
Jänicke <i>et al.</i> [43]	The Bible translations	-	24	Variant graphs	English
Geng <i>et al.</i> [26]	Literature (Othello)	10	10	Parallel coordinates, heat maps, scatter plots	German
Riehmman <i>et al.</i> [64]	PhD theses and Wikis documents	2	Not-specified	Slope graph, glyph-based visualization	English
Asokarajan <i>et al.</i> [5]	Classical Latin texts	-	22	Multiple views+pixel-based visualization, dot plot	English
Cheesman <i>et al.</i> [21]	Literature (Othello)	1	2, 40	Juxtaposed text versions, stylometrics diagram	German
Silvia <i>et al.</i> [80]	Classical Latin texts	-	12	Storyline visualization	Latin
Jänicke <i>et al.</i> [49]	Medieval texts	2	2	Juxtaposed text versions	French
Abdul-Rahman <i>et al.</i> [2]	18th-century literature	2	2	Parallel Coordinates, dot plot	French

TABLE 1

Summary of visual design characteristics for related work discussed in the sections: 3.3, 3.4 and 3.5. The “-” sign in the table indicates that an approach does not provide a distant or a close reading view. “Arbitrary” means the number is open based on the author’s claims and “Not-specified” means the number of parallel documents is not mentioned in the paper and is not exemplified. The references are ordered based on publication date.

[74] which enables the user to investigate multiple documents side-by-side and integrates linking functionality to highlight corresponding text fragments. Similarly, multiple approaches use a side-by-side layout to represent compared objects, such as Jong *et al.* [33], Welsh and Hooper [84], Behrisch *et al.* [8], Wheelles and Jensen [86], the text view by Geng *et al.* [26], the text reader by Jänicke *et al.* [48], Howell *et al.* [38], Cheesman *et al.* [21], and Jänicke *et al.* [49]. Asokarajan *et al.* [5] visualize the variation in a pixel-base matrix where the x-axis represents the offset in the text and the y-axis represents the variation (witnesses). They also visualize the summary of variation at the pages, lines, and words level. We extend the discussion of some of these approaches in Section 3.4.

Plagiarism detection is an application of visual text comparison as in White and Joy [87]. Also, Riehmman *et al.* [64] combine an overview slope graph and glyph-based detailed representations to explore given text against multiple sources.

Different approaches overlay parallel texts in the same coordinate system in order to communicate comparative objectives. For example, the variant graph in Jänicke *et al.* [41], Storylines in Silvia *et al.* [80], and Geng *et al.* [26], [27]. The variant graphs and parallel coordinates visualization represent each object as a line in the visual space. In the variant graphs, the y-axis illustrates the offset in the text or time.

There are approaches which extract relationships between parallel documents and explicitly visualize them to support visual comparison. The Stylometric representation of versions in Cheesman *et al.* [21] encodes the similarity between connected texts using the thickness and length of the

links. Explicit encoding of relationships can be implemented using a dot plots representation to detect similarity and dissimilarity patterns, such as in Ribler and Abrams [63], in Jänicke *et al.* [48], and in Abdul-Rahman *et al.* [2]. Collins *et al.* (Parallel Tag Clouds) [22] also uses links and word clouds to encode the relationships between documents.

The system we present uses juxtaposition but with up to 38 close as well as distant reading of parallel translations. It also features explicit alignment curves. The exploration and interaction techniques are customized and implemented to satisfy the user requirements and tasks stated in Section 4.

3.4 Examples of Visualizations from Digital Humanities

There are many visualization solutions that enable digital humanities’ primary tasks. In this section, we focus on visual approaches that feature alignments between parallel texts. Jong *et al.* [33] present an interactive tool that conveys the structure of parallel texts using color-coded boxes representing words. The reader can switch between the structural view and the textual view to facilitate close reading.

Büchler *et al.* [16] introduce a graph visualization to illustrate citation variation among documents in a corpus. They provide a distant reading view of the citation in the corpus using an interactive bar chart.

Howell *et al.* [38] propose a close reading visual design driven by digital humanities methodologies of the novel *The Secret Scripture* (2008). They visually compare two different encodings of the same novel using different color-coded highlights.

Jänicke *et al.* [48] introduce distinctive contributions to this field. They introduce multiple visual designs to depict text re-use between collections of text. For distant reading,

they design a visual matrix to discover the type and amount of text re-use between pairs of texts. Additionally, they introduce the text re-use reader which consists of two panels: a dot plot view and a text reader. The former view depicts the type of text re-use between two texts, e.g. a diagonal pattern indicates sections repetition while vertical and horizontal patterns indicate phrase re-use. The text reader shows two panels aligning two documents, both panels are linked and respond to one another. They also introduce the text variant graph [42] to detect variations between versions at the sentence level. The graph uses color-coded links for each version and font size to encode the number of occurrences among all versions. They demonstrate a graph applied to five versions of the Bible.

Jänicke *et al.* [43] propose an extended distant view of the variant graph to support analysis on higher text abstractions such as sections or chapters. They exemplify their method with a distant reading of 24 Bible editions.

Jänicke and Wrisley [49] propose a visual analytics environment that supports aligning two versions, or more, computationally. Also, the tool integrates different interactive methods that analyze textual alignments, along with an intermediate view between close and distant reading which they call “Meso reading”. The Meso reading view combines the text and the statistical features together within the same visual field.

The difference between our work and the related work discussed in this section is that our proposed design connects a distant reading view of all 38 translations with the close reading view in a novel way by smooth zooming and panning. Interactive zooming and aligning facilitate comparison of the related speeches across a number of parallel texts (Section 5). Previous work separates the close and distant reading views in multiple windows. The user is required to cognitively integrate the two. Also, we encode means to help the user validate alignment or translation of the segments using similarity metrics. The Eddy and Viv metrics are calculated interactively and dynamically to reflect the similarity among the current selection of parallel translations. Among all of the related work, our proposed design deals with a cross-language dataset and presents a macro (distant) view of the entire collection. Previous work does not generally support comparison of over 30 aligned translations and is generally restricted to the English language or a single language.

In Section 3.5, we introduce the related literature that analyses and provides visual designs of Shakespeare’s *Othello* collection.

3.5 Previous Work on Shakespeare’s *Othello*

In this section, we discuss previous visualizations of Shakespeare’s *Othello*. Geng *et al.* [27] introduce a focus+context parallel coordinates layout for comparing eight translations of Shakespeare’s *Othello*. Their design consists of two main components. The first is a distant view represented by parallel coordinates to show the variation between translations and the use of the most frequent words. A collective concordance of the most frequent words is shown in the column on the far left. Each coordinate represents a word-translation pair, and the thickness of the bar encodes the similarity rank. They support various interaction techniques to aid

exploration and investigation, such as brushing, selection, and linking. The second component of the visual design is the close view which shows the actual text and highlights the selected keywords.

Geng *et al.* [26] integrate multiple visual designs to illustrate the similarity between subsets of translations of Shakespeare’s play *Othello*. They provide visual designs to support distant reading, such as heat maps to illustrate segment structure, and parallel coordinates to depict similarity among versions. In the text view, close reading is obtained by showing the text segments in multiple versions of the play.

Cheesman *et al.* [21] present a web-based tool that enables the user to create parallel, segment-aligned multi-version corpora. The main goal of their project is to digitally explore patterns of variation among multiple translations. They present two overview designs which provide a distant reading view of the corpus. A small multiples pairwise alignment map of 35 German translations is used. Each translation is aligned with the base English text. Each speech is represented by a vertical rectangle and the height of the rectangle encodes the length of the speech. The edges between each translation and the base text represent alignments between segments. The viewer can identify different attributes of each text and make comparisons, such as the variation in length between translations and the base text. In the same overview context, they provide more analytical and statistical aggregations of the corpus data represented by a stylometric network. The network diagram shows translation clusters which depict the similarities among versions. The connection edges represent the similarities in particular sets of frequency counts, and the thickness of the edges reflects the degree of similarity.

Additionally, Cheesman *et al.* [21] provide a detailed interface which aligns segments of the base text with translated versions along with similarities metrics (Eddy and Viv) (Explained in Section 2).

The difference between the work presented in this section and our work is that Geng *et al.* [26], [27] present only a subset of the collection and is difficult to scale accordingly. On the other hand, Cheesman *et al.*’s [21] alignment map aligns only one translation with the base text and does not encode any similarity features. The Eddy and Viv interface provides only a close reading and does not allow filtering and selection of translations. Our work supports the comparison of the whole collection and incorporates the encoding of similarity metrics. It provides a variety of interaction and exploration techniques that facilitate the analysis and visualization of the collection. There is some overlap of co-authorship between this current work and previous work on Shakespeare’s *Othello*. We exploit the previous studies to guide the current work.

In summary, the current work is unique in that it includes 38 translations along with the source text. The current work supports integrated distant and close reading in the same view and implements them with smooth zooming and panning. Also, it allows the user to explore different regions of interest and stores them for further analysis. The current design allows the user to interactively customize the alignment overview, and subsequently update the similarity metrics based on the user’s preference. Although, the TLC

(Term Level Comparison) design is not novel, including it in the process of exploring the dataset is novel and proves useful. See Sections: 5.1.3 and 6.2 for practical use of the TLC view.

4 DESIGN REQUIREMENTS AND TASKS

The original question that was posed for the application to address is: how can the variation between any number of parallel translations of a given source text be represented visually so as to enable users (a) to identify overlaps, absences, additions, and variation between parallel translations and (b) to study the findings of various kinds of algorithmic, comparative text analyses. So, the rationale behind our visual design is to enable users to interactively explore the translation collection to answer this question. To achieve this, we established and incrementally refined a list of requirements. The requirements that our proposed implementation fulfills are as follows:

- R1** An application that enables comparison of translated parallel text.
- R2** A visual design that supports both close and distant reading.
- R3** A layout that supports close reading for further detailed analysis.
- R4** A visual design that considers stable versus unstable translations.
- R5** Interaction that enables customization of the translated texts.
- R6** Interaction that supports general exploration and analysis.

To action this list of requirements, we established a list of associated tasks for implementation. We derived five main tasks based on the typology of visualization tasks by Brehmer and Munzner [14] in order to achieve the aforementioned requirements and motivate our visual design:

- T1** In this *discovery* process, the user should be enabled to *explore* the alignment overview in order to *identify* a region of interest (**R1**). [discover→explore→identify]
- T2** After identifying a region of interest (**T1**), the user may *navigate* the space leveraging smooth zooming and panning. As the user navigates, multiple levels of details are *aggregated* and rendered (**R2**, **R3**, **R6**). [T1→navigate→aggregate]
- T3** As the user performs **T1** and/or **T2**, the user may apply different *filtering* and *selection* tasks to assess the exploration task (**R5**). [T1/T2→filter/select]
- T4** As the user performs **T2**, the user may *select* a segment to obtain details-on-demand (i.e. a close reading view) (**R3**). [T2→select]
- T5** As the user performs **T1**, **T2** and/or **T4**, the user may perform interactive *comparisons* of the parallel translations exploiting meta-data based on the alignment of speeches and text similarity metrics (**R4**, **R6**). [T1/T2/T4→compare]

We relate to these tasks in the discussion of our proposed design (Section 5). Our design is influenced by the visual information seeking mantra by Shneiderman [78] that suggests providing an overview first, then zooming and filtering options, and finally details-on-demand. The design is also influenced by previous work on this topic which aims

to align texts side-by-side to support comparison tasks. Additionally, the design is also guided by careful collaboration with the domain expert.

5 TRANSVIS'S DESIGN

In this section, we introduce our proposed interactive visual design of parallel translations and relate our choices to the tasks from Section 4. Our system is composed of four main constituents starting with an overview.

The first window offers the **alignment overview** of parallel translations of Shakespeare's *Othello* (**T1**, **T5**). It provides a general context for understanding the collection and conveys the whole dataset in one visual layout. Furthermore, it leverages interactive capabilities to enable the user to explore and find interesting patterns and features within the collection. The alignment overview allows users to examine significant, larger patterns in the translations which are not readily viewable from narrow or detailed views. Window (A) in Figure 1 shows the distant reading view of 38 parallel translations aligned with the base text. The curved edges between translations depict alignments between speeches. The zoomed-in portion in Figure 1 shows a close view of the curved edges and segments.

Design justification of the alignment overview: Our data is high-dimensional. We present 38 texts and each text contains multiple levels of abstraction (term, segment, speech, and books). Thus, we present a parallel view of translations and in each translation we present the encapsulated structure. Juxtaposition supports visual comparison of alignments intuitively, when comparing different manuscripts the user places them spatially next to each other and performs comparison. Also, we incorporate a number of exploration techniques to help the user customize the alignment overview to their preference such as by zooming, filtering, or selecting.

The second main component is the **options panel**, shown in Figure 1 (B). It provides the user with a range of layout functions in order to facilitate exploration of the collection and comparison of translations (**T2**, **T5**). The user can perform a query-based search and the results are visualized using focus+context in the main window (A). It features a number of tabs that support different tasks. The first tab is for color properties which enables the user to modify the color mapping schemes and color-map different properties of the speeches, such as individual speech length and language similarity values. The second tab is the options tab where the user may alter the properties of the visual design, such as the order of the translations and length of time tooltips are shown over speeches. The third tab is the filter tab. Different filters are provided for the user to reduce the complexity of the visual design, such as stage direction and speaker filters. These filters are discussed in Section 5.2.3. The last tab is the time-oriented thumbnail view. In this section, all of the translations are depicted using color-coded thumbnails. If the translation is shown in the main window (A), the thumbnail is green otherwise it is red. These filtering and selection options are the result of our feedback sessions with the domain expert. This is described in more detail in Section 6.1.

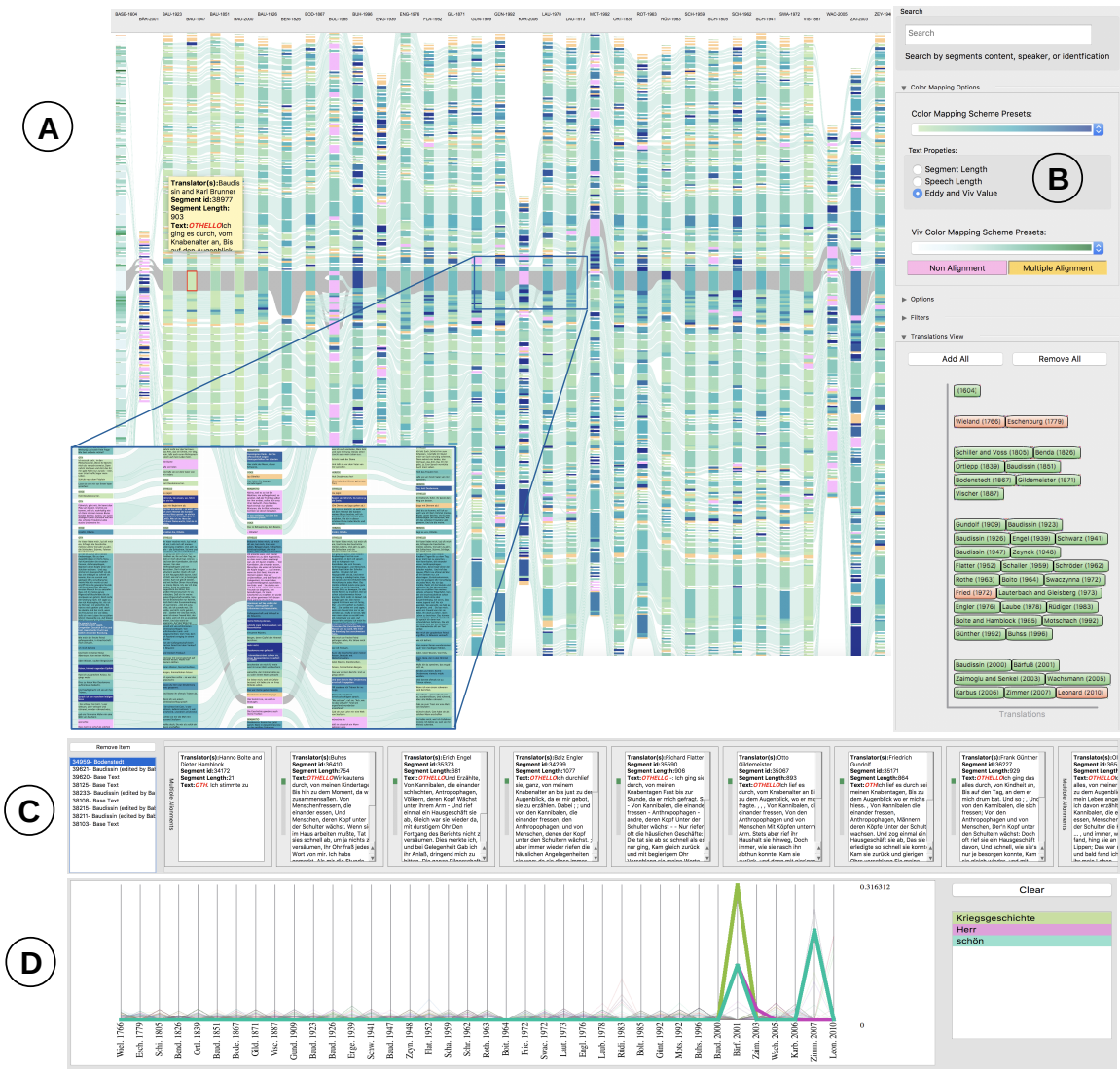


Fig. 1. The alignment overview (A) shows the parallel alignment of translations with the original base text. The highlighted path in (A) shows the distant alignments of a segment of the "Othello" speech starting with "I ran it through...". In the left-bottom, a zoomed-in view magnifies the curved edges. Window (B) shows the options panel that facilitate exploration of the collection and comparison of translations. View (C) is a close reading view (the detailed view) that corresponds to a user-selected speech and each aligned speech. Window (D) shows the Term-Level Comparison (TLC) view. The zoomed-in rectangle is part of the figure, not of the visualization itself.

Design justification of time-oriented thumbnail view: Most of the user options are implemented in close collaboration with the domain expert. The thumbnail view clusters the translations chronologically which makes it intuitive and quick to explore and navigate. Adding and removing translations is simply performed either by toggling a translation on or off or dragging and dropping. It is important to let the user customize the starting point when there are so many translations. The view aids the user by visually informing him of the selected translations (green buttons) and deselected ones (red buttons).

The third main component of the system is the **detailed view** which is the focus subset that the user is interested in after performing filtering and selection (T4). As shown in Figure 1 (C), the detailed view shows a close reading of the user-selected speech along with the aligned speech translations. It stores both the interesting aligned segments (path) and the alignment overview state, so the user is able

to revisit any previously selected paths for further analysis (T5). Also, the text in this view is accessible to user and can be copied to be used beyond our system. The detailed view is discussed in Section 5.1.2.

Design justification of the detailed view: It is recommended to allow the user to have access to the actual text particularly when developing visualization for literary scholars [33], [47]. This close reading gives complete access to the text for further analysis with other software. This view aligns the segments in a compact and simple context for further analysis. It incorporates a list of previously selected paths to facilitate comparison between paths and also to enhance the user experience by saving the actions history (T5). We provide this addition of close reading to support close and distant reading simultaneously without losing the user's first (or previous) choice of speech for close reading.

The fourth component is the **Term-Level Comparison (TLC)** view (Figure 1 (D)). The TLC is an interactive ana-

lytical tool that assigns weights to each word in each user-chosen segment aligned with a base English segment. We use TF-IDF as a term-weighting to measure the significance of each original word or lemma in a segment with respect to the word occurrence in the whole corpus (the segments aligned with the base English text in our case). The view clearly justifies the Eddy metrics and signifies the terms (original or lemmas) that define segments. Also, it aids the user in finding terms and translation variance (T5). To show this plot, the user can drag an interesting path, usually because of an Eddy value distinctiveness pattern, then drops it onto the TLC view area to see the word's contribution in this path. This view is also motivated by the on-going discussion with the domain expert requesting close analysis and details.

Design justification of the TLC view: As TLC view could be considered as parallel coordinates, it presents strength when exploring and processing multi-variate data [40]. Such techniques are useful to explore anomalies in the data even without an extended outlier-detecting mechanism [60]. As a result of the TF-IDF, the commonly used terms are assigned lower values and vice versa. Thus, the distinctive terms will stand out very clearly in the TLC view which is the main motivation behind this design. To overcome the limitation of the view when there are cluttered lines, the view provides a list of words in alphabetical order. When the user selects any word, the TLC highlights the line corresponding to this word. The user also can enable brushing to highlight multiple lines. The color opacity of the unselected terms is reduced in order to remain as context. The terms list updates to reflect the brushed terms. The terms list also assigns the same color of the term's line to the list item to visually identify the line's correspondence as shown in Figure 3. The colors of the lines are automatically generated to uniquely assign a color to each term.

5.1 Visual Design Factors

We separate the primarily visual and interactive design features to facilitate reading.

5.1.1 Filtering based on Derived Alignments

The dataset contains alignment information between each speech in each translation and the English base text. However, it is useful to have an alignment between any two arbitrary translations. We derive meta-data that aligns two arbitrary translations with respect to the base text. We align two speeches from different translations if they correspond to the same base English speech.

With the alignment meta-data, we have the alignment between the source text segments and the translations' segments. The derived alignment is the alignment between a translation and another translation. If the translation is not adjacent to the source text, we derive the alignment between two segments (a and b) of two translations (A and B) if b has an actual alignment and a has an actual alignment, we connect them. The path of alignment disconnects if one of these conditions break. For instance, if b has an actual alignment with the source segment, and a which belongs to translation A that occurs before translation B does not have an actual alignment, the path of alignment disconnects. This mechanism is implemented as a user option, and as

explained it results in filtered alignments as the path moves left-to-right. However, the user-chosen order of translations affects the results in the alignment overview and the user needs to choose the sorting function carefully.

This mechanism can be useful, particularly when comparing translations and as a filtering option (T3, T5). However, in some cases, it might not yield results when the translations are not related, such as when translations stem from a different era or author.

5.1.2 Detailed View For Close Reading

In addition to smooth zooming in the alignment overview that enables the user compare specific speeches, the user is able to select any speech to analyze and compare it along with the aligned speeches in a dedicated detailed view as shown in Figure 2 (T4, T5). The detailed view provides another close reading option for the speeches such as translators names, speech identifier and Eddy value (discussed in Section 2). The close view shows the user-selected segment using a highlighted red border and each aligned speech in a scrollable window that is easy to read and investigate (T5). Each speech is paired with a colored bar showing the similarity distance to indicate the distinctiveness between aligned speeches. The longer the bar, the more distinctive the speech is with respect to the other translations.

In order to improve the user experience and to ease the exploration and analysis task for the researchers, we maintain a history of all previously investigated speeches in a list as shown in Figure 2 (bottom-left corner). When the user revisits any of the previously examined speeches in the list, the corresponding speech and the edges of alignments are highlighted in the alignment overview. The interactive history list identifies each user-selected alignment using the segment identification and the translation name such that the user may remember their own user provenance with respect to the alignment selection.

5.1.3 Term-Level Comparisons (TLC) View

The TLC View illustrates the weighting of each term using line charts. The y axis shows the normalized weighting of the terms. Along the x axis, we render vertical lines to indicate the translations. In this view, we can explore each term weight across all translations. The TLC view (Figure 3) facilitates comparisons and analysis tasks (T5). For the term-weighting, as discussed in Section 2, we use the TF-IDF weight for each word.

Alignment overview with TLC view: The rationale of this view is to observe and explore the aligned segments in the term level and highlight outliers and uncommon terms. The TLC shows the terms that contribute to the pattern discovered in the alignment view when using the similarity metrics.

To explore aligned segments, the user can drag-and-drop any segment of the path from the alignment overview into the TLC view. A list of the terms is also presented on the right. If the user selects any word in the list, the corresponding line is highlighted, and vice-versa.

In Figure 3, three distinctive terms are highlighted in the TLC view: "Neger", "Schwarze", and "Maure". These translations correspond to the word "Moor" in the context: "I hate the Moor." These terms are illustrated in the three brushed peaks shown in the figure. The terms list shows the

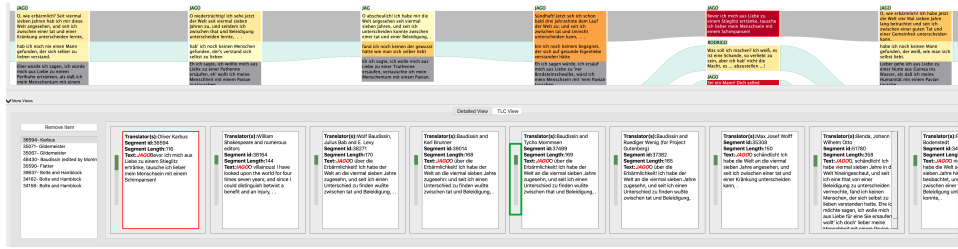


Fig. 2. The detailed view shows the user-selected speech highlighted using a red border and the aligned speeches ordered consistent with the translations appearing in the alignment overview. Each speech is paired with with a colored bar (annotated with a green border) to indicate the similarity distance. In the bottom-left corner, a list of all previously selected speeches. If the user selects any speech from the list, the corresponding text is highlighted in the alignment overview and the edges of alignments are presented above.

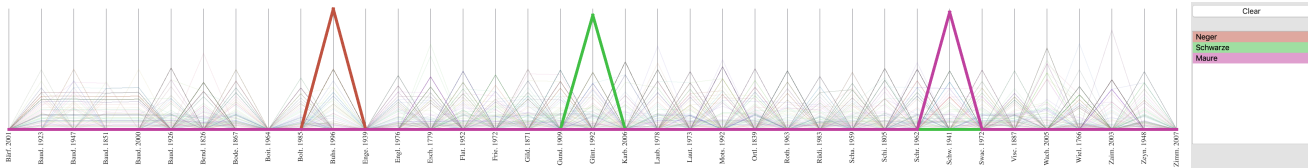


Fig. 3. An example of aligned segments of the base English segment starting with “*Thou art sure of me...*” depicted by term-weighting matrices using the TLC view. In this example, the user brushes the three peaks that reflect the distinctive translations of the word “*Moor*” in the context: “*I hate the Moor*”. The highlighted translations are “*Neger*” in Buhss (1996) [17], “*Schwarze*” in Günther (1992) [32], and “*Maure*” in Schwarz (1941) [76]. The terms list reflects the brushing result and assigns the same colors to the terms. The user-chosen terms are rendered in the focus while the rest are rendered as context.

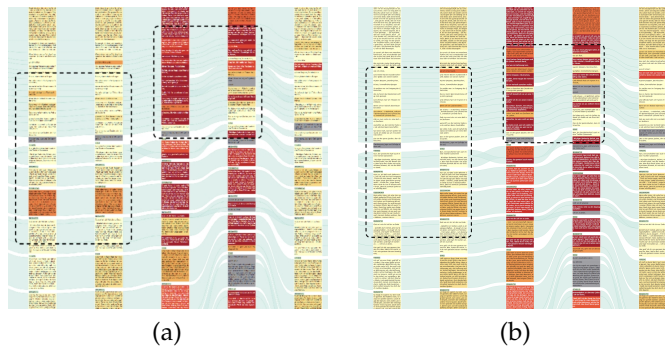


Fig. 4. (a) An example of two corresponding sub-sets of six editions of Baudissin (1832) [7]. The variation of colors in the dashed rectangle in the original version (a) is clearer than the corresponding rectangles in the lemmatized version (b).

only selected terms and the color facilitates identifying the correspondence between the lines and the terms. This selected terms are rendered in focus and the rest are rendered in context. See the domain expert feedback in Section 6.1 for further discussion on this.

Lemmatization (as discussed in Section 2) is a normalization algorithm used to reduce inflected words and identify the base word [52]. We use CAB [45] to create a dictionary of the lemmas for our corpus. Processing such an antiquated language is challenging due to many missing terms.

Example-lemmatized versus original text: We compare the results between a lemmatized and non-lemmatized version of the translations. Our corpus includes several editions of the classic, canonical Baudissin translation, first published in 1832 and often re-published (100s of times so far): printed editions from 1851, 1923, 1926, and 1947, and a digital edition from 2000. In Figure 4, we can see in the original version (a) that the variation of colors in the dashed rectangles is more distinctive than the corresponding areas in the rectangles in the lemmatized version (b). The variation in

(a) is clearly a result of inflected words that share the same lemma.

Although, technically, TLC view is not a parallel coordinates view, it inherits some of its limitations. The TLC view can be difficult to interpret and explore in the case of long segments due to over-plotting. Much research focuses on interaction techniques to overcome the visual clutter challenge in parallel coordinates plots by reducing the dataset or by reducing or modifying the order of the dimensions [44]. In our case, the combination of TF-IDF as a term-weighting metric causes the common terms to clutter in the areas of low x values. However, as this can be problematic when searching for common terms, it helps to identify rare vocabulary in each translation.

The alphabetically-ordered word list that accompanies the view can also help the user find term line by clicking on a word and highlighting the corresponding line. The stop words removal, or lemmatization can make the plot less cluttered.

5.2 Interaction Design Factors

In this section, the primarily interaction design factors that our system encompasses are elaborated. The design is heavily based on user preference and interaction to customize the output. In general, the interaction starts with the alignment overview, the system by default shows the entire collection. The user can interact with the system to modify the alignment overview using the options panel, shown in Figure 1 (B). For example, the user can change the order of translations based on a range of sorting presets, filter the alignment overview based on a speaker, or change the presented translation using the time-oriented thumbnail view (Section 5.2.3). Then, the user can use the smooth zooming and panning (Section 5.2.1) to explore and identify passages of interests. Once the user finds an interesting pattern guided by the predefined customization and the similarity-based color-coded segments, the user can use the detailed view

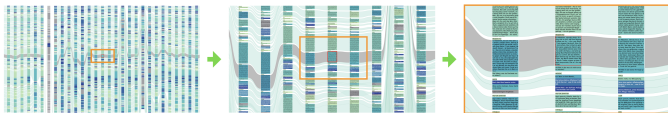


Fig. 5. An example of three levels of integrated zooming and the smooth changes to the level of detail. (left) Distant reading without zooming. (middle and right) close reading after the user zooms in. The textual content of the speeches fades in smoothly at an increasing level of detail. The grey curved path indicates a user-selected alignment.

(Section 5.1.2) to explore or save the path of alignment by clicking on it, or dragging the path and dropping it on the TLC view (Section 5.1.3) to examine the terms contained in the path. When the user clicks on individual segments, the system saves them and allows the user to revisit them and retrieve the alignment overview. Both of the case studies (Section 6.2) demonstrate the interactive overall process that facilitates the finding.

5.2.1 Smooth Zooming of Translations

In our design, we start with the alignment overview of the whole collection of translations. The user is able to zoom in smoothly and fluidly to explore the dataset and investigate (T1, T2). When zooming in smoothly, more details fade in gradually, such as the actual text content of each speech to support close reading. See Figure 5.

With zooming as an option, the user is able to investigate a region of interest in the same view and has the ability to transition between close and distant reading without switching between multiple windows. When zoomed in, the user is able to pan smoothly for comparison of translations (T4, T5).

During the zooming, we maintain certain levels of detail as shown in Figure 6 in the supplementary file. The first level illustrates the speech showing the length of the segments. Also, the relative thickness of the border highlighting the user-selected segment is increased relatively. We decrease the thickness of the highlighting border as the user zooms in. In the second level, more segment details are revealed, such as local colors and the length of the text. Finally, at the third level, the text is readable and the thickness of the highlighting border is decreased considerably. This functionality is implemented in the alignment overview and it zooms along both x and y axes with respect to the mouse position. Zooming in on only y axis could cause only a single word to appear on each line which is difficult to read. Zooming in on x axis could result in long, difficult to read lines and difficulty comparing alignments.

Multiple works link different abstractions of single text using zoomable layouts. VarifocalReader [51] uses a combination of x and y axes zooming to show the multiple layers of abstraction. However, it becomes challenging when comparing more documents, and the number of abstractions layers needs to be decreased. Gold *et al.* [30] integrates the close and distant reading of a single text using zooming along the y -axis. On the other hand, Asokarajan *et al.* [5] use the zooming along the x axis to examine the view closely and do not provide a level of abstraction based on the zooming.



Fig. 6. Two snapshots of the same region of the collection. In (a) the segments aligned with the user-selected segment are out of view and not visible. In (b) the segments are horizontally aligned with the user-selected segment.

There are other approaches that utilize interactive lenses [18], [65]. We believe that magic lens and interactive magnification techniques that allow the user to obtain a close picture of the interested area could be useful. However, they also introduce new drawbacks such as either distortion or discontinuity between the focus and context areas. We discuss these techniques with the domain expert and believe such alternatives could be explored in future through task- and design-driven studies. We did not incorporate them because the user requirements are fulfilled with our current design choices. This discussion also leads to the identification of an important pitfall in design studies. See Section 7 for details.

We compute the average time (milliseconds) of rendering the scene while performing the zooming. For each number of translations, we record the average time for the process. See Figure 7 in the supplementary file. Our system takes about 5.6 ms to render the scene. To demonstrate the scalability and the linearity between the number of translations and performance time, we modified the dataset to increase the number of texts to 76 translations and 1 base English text. We can see from the figure that the time in general increases as the number of translation increases. We perform this experiment using a machine with the following specification: Processor: 3.3 GHz Intel Core i5, operating system: Mac OS version 10.14, memory: 8 GB DDR3, and an AMD Radeon graphics card with 2 GB of memory.

5.2.2 Filtering, Selection, and Positioning in the Alignment Overview

Filtering and selection aim to reduce the complexity of the scene by abstracting some elements of the translations to help the user focus and find regions of interest. The user can click on an individual speech and see the corresponding translated and aligned speeches interactively. Within each translation, horizontal bars represent a speech or a segment depending on the user preference. The vertical order of bars implicates the position of the corresponding speech or segment in the text. The thickness illustrates the length of the speech or segment calculated in words. However, when adjacent translations are distant from one another vertically, it may become difficult to compare the aligned speeches, specifically in zoomed-in views, as illustrated in Figure 6 (a). Thus, the user can choose to interactively align them vertically side-by-side (T5), as shown in Figure 6 (b).

We connect the corresponding segments using curved edges as illustrated in Figure 6. The user-selected alignment is highlighted to be distinct from other alignments. The user

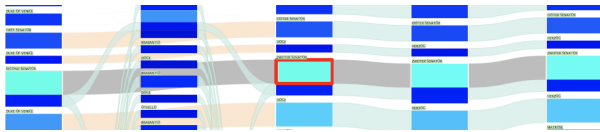


Fig. 7. The placement of the unstable translation by Bärffuss (2001) [6] results in multiple disconnected edges between non-adjacent translation. The differently colored edges illustrate the alignment between the two non-adjacent translations.

may also smoothly re-arrange the order of the translations manually through a drag-and-drop mouse movement. In our feedback and evaluation sessions with domain expert from the Arts and Humanities, this feature was used extensively.

Example–Juxtapositioning Aligned Speeches: We implement a mechanism to smoothly translate corresponding speeches vertically as seen in Figure 6. In Figure 6 (a), the user-chosen segment and the aligned segments are not in the alignment overview as a result of their original placement within the translation sequence. In Figure 6 (b) the translations slide vertically to line up and with the user-chosen segment. Thus, the user is able to compare and explore the segments in the same close view (T5).

A configuration can occur, if two adjacent translations are not aligned. This might lead to confusion. To address this limitation, we implement a user interaction technique to render such edges that connect non-adjacent translations using a different color and rendering order as shown in Figure 7 which illustrates the problem when examining an unstable translation such as Bärffuss (2001) [6]. However, this is still can be challenging particularly when dealing with unstable translations.

5.2.3 Filtering and Selection of Speeches and Translations in Options Panel

In the options panel (Figure 1 (B)), the user is able to filter and compare the translations in a variety of different ways (T3, T5). We found that this option was always the first user-option chosen by the domain expert (see case studies in Section 6.2).

One option is filtering the data based on a custom query. The user is able to search for a given speaker, a specific segment ID or word. Also, the collection can be filtered based on a speaker name if the user is interested in a specific character. The speaker search is non-trivial since the speakers are translated differently from one translation to another, e.g. the speaker “Duke of Venice” is translated for example into: “Der Doge”, “Herzog”, and “Doge” and the speaker “First Senator” is translated into: “1 Senator”, “Senator” and more commonly “Erster Senator”. However, to address this issue and to reveal more relevant results across both the base English text and the translations, we populate the speaker list with all of the characters in the original text along with those not aligned with the original text. When the user looks up a speaker, we can eliminate all characters not aligned with it. Thus, we are able to depict all translated speakers in the translations. See Section 6.1 for the domain expert feedback.

Example–Filtering Based on Speaker: In some segments, we are able to see transformation in speaker names, e.g. the speaker in the speech: “Duke of Venice: Dead?”

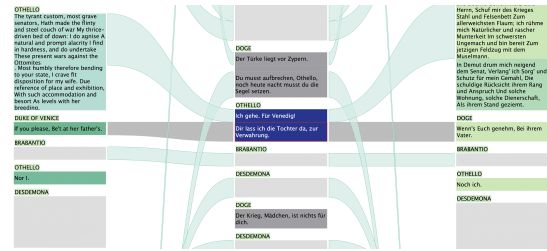


Fig. 8. A subset of a filtered focus+context rendering of alignment overview. The view is filtered by the speaker “Duke of Venice”. The high number of matches is due to mistaken correspondence during the original segment alignment process which is easily discovered by the visualization result. In this figure, a segment of speech by “Othello” was mistakenly aligned with a segment of speech by “Duke of Venice”.

is transformed differently into “Erster Senator: Tot?” as in Flatter (1952) [25], “Die Senatoren: Tot?” as in Zeynek (1948) [91] and “Senator: Tot?” as in Wachsmann (2005) [83]. In some cases, speaker transformation is accepted since the source text might indicate different speakers for the same speech and translators interpret this with a variation.

However, in other cases, the process of the alignment is not accurate and the speaker filter can facilitate the discovery of errors during the original, labor-intensive alignment process. For example, in Figure 8, the “Duke of Venice” speaker is selected to show only the corresponding segments. However, a great volume of matching translated segments are connected to our surprise (≈ 1500 segments). After investigating, we find that the aligner has mistakenly linked a speech by “Othello” to the speaker “Duke of Venice” which causes this volume of unexpected results.

In all of the above filters, the results are shown using focus+context. The results are visually highlighted while the context is preserved in greyscale.

Some of the segments in the collection are not spoken in the play but provide directions to the characters. These segments are called stage directions and can be toggled on or off (T3, T5). Most of the time, researchers are not interested in studying them and removing them can reduce clutter from the scene.

The researcher is able to use the thumbnail view tab to **add or remove translations**. Adding and removing translations is achieved using a smooth drag and drop interaction or simply by selecting and deselecting translations (T5). The translations in the options panel as seen in Figure 1 (B) are color-coded green or red according to the translation presence or absence in the alignment overview (A) respectively. See case studies Section 6.2 for an example of this used in practice.

Additionally, we implement various **translation ordering and sorting** options to aid researchers in finding the best layout that supports comparison of translations (T5). The options provide an initial layout that suits the researcher’s needs. Some of the options have been suggested by the domain expert, such as sorting the translations chronologically. Other options sort based on the aggregation of the similarity metrics, e.g. average and total of Eddy values. See case studies Section 6.2 for an example of this used in practice.

The proposed visual design implements different **color mapping schemes** to assist researchers in investigating the

parallel translations. There are different primary color mapping schemes as shown in the accompanying video. In the center, there are three translations attributes that we color code the text based on (segment length, speech length, and Eddy and Viv value). In the bottom section, we provide the user with a range of color mapping scheme presets for the Viv values. The user is able to customize specific translation attributes which are multiple-aligned segments and non-aligned segments. See Section 6.1 for a case study demonstrating the utility of these color-mapping schemes. The sequential color schemes are generated using Color Brewer [34]. The rainbow-style color scheme is generated using Telea’s algorithm [82].

6 EVALUATION

We work closely with the domain expert and the project is driven by a real-world historical investigation. See Figure 10 in the supplementary file. With the close observation by the domain expert, we designed the application to satisfy the user requirements in Section 4 and facilitate exploration and interaction to enhance the user experience.

The domain expert feedback demonstrated in Section 6.1 is based on the 9 in-person, regular sessions to demonstrate the integrated visual design features over 5 months. All of the sessions are video-recorded for post-analysis and archiving gathering. Our semi-structured interview questions were previously planned and guided by Hogan *et al.* [37]. The total number of hours of feedback is around 9 hours. The first few sessions consisted mostly of software demonstrations to guide the development of features. See Figure 10 in the supplementary file. However, these gradually turned into an active hands-on use of the software by the domain expert. The software also evolved due to feature requests. We implemented features that give researchers more control over the exploration, comparison, and analysis of parallel translations. During the face-to-face feedback sessions, there are different patterns that were observed, such as, the discovery of software bugs, and the discovery of data-level errors.

In Section 6.2, we present two case studies conducted to validate our visual design and the integrated similarity metrics.

Domain Expertise: The domain expert is Professor Tom Cheesman, in the Department of Modern Languages, Translation and Interpreting in the College of Arts and Humanities, Swansea University. He is the principal investigator on the “Version Variation Visualisation” project. The project is responsible for collecting, aligning, and warehousing the dataset that we are examining, and other ‘multi-translation’ datasets and the team has developed prototype online tools [19] for managing such datasets and visualising comparative analyses of them. Professor Cheesman is a specialist in modern and contemporary German literature and culture. He has been researching German culture and translating German literature since the early 1980s. Professor Cheesman has been investigating the history of German translations of Shakespeare’s play *Othello* since 2009, using both traditional qualitative methods (contextualised close reading) and experimental, quantitative, digital methods.

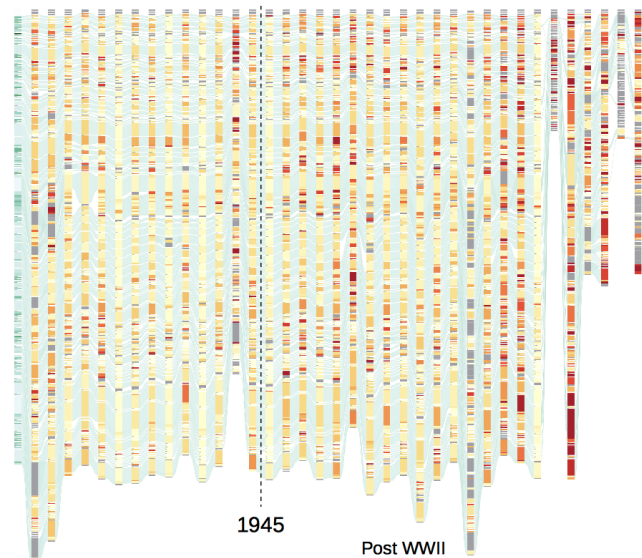


Fig. 9. The alignment overview of the translation collection reveals an increase of variation between translations particularly after the second world war, with the exception of Engel (1939) [24].

Relevant online outputs, presentations, and published articles by him and his collaborators are listed on the project’s website [19]. The articles include publications in *Digital Scholarship in the Humanities* [21] and *Journal of Data Mining and Digital Humanities* [67].

6.1 Domain Expert Feedback

The Alignment Overview Feedback: The alignment overview enables the researcher to capture global patterns and direct attention to a region of interest for a more detailed investigation. When demonstrating the camera-positioning features in the alignment overview, the domain expert states, “In a birds eye view, the visual design shows how versions differ in length, and the number of alignments, and the Eddy color mapping neatly highlights (a) versions which are generally a high-Eddy and (b) segments and passages in the run of text which are high-Eddy. This is great.”

The color mapping used to illustrate different translation attributes in combination with sorting options may reveal new insight to the domain expert. As seen in Figure 9, the domain expert notices that the variation between translations increases distinctively after the second world war. Also, he stated that, “modern translators increasingly diverge from the norms of theatrical Shakespeare language established in the 19th century and early 20th versions.”

Search Feedback: The focus+context rendering of alignment overview of search results can help the user

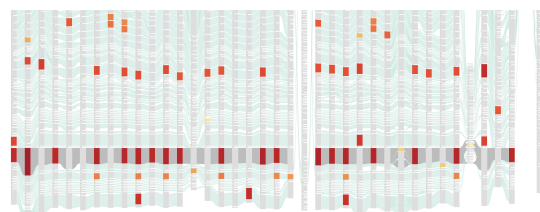


Fig. 10. Focus+context rendering of alignment overview of the results of the search for the word “Lust”.

Bau. 1851	Bau. 1923	Bau. 1926	Bau. 1947	Bau. 2000
<p>JAGO</p> <p>Nun, wenn du das tust, so ist's mit meiner Freundschaft auf ewig aus. Ei, du alterner junger Herr!</p>	<p>JAGO</p> <p>Nun, wenn du das tust, so ist's mit meiner Freundschaft auf ewig aus. Ei, du alterner junger Herr!</p>	<p>JAGO</p> <p>Wenn du das tust, ist's aus mit meiner Liebe, du alterner Kerl!</p>	<p>JAGO</p> <p>Nun, wenn du das tust, so ist's mit meiner Freundschaft auf ewig aus. Ei, du alterner junger Herr!</p>	<p>JAGO</p> <p>Nun, wenn du das tust, so ist's mit meiner Freundschaft auf ewig aus. Ei, du alterner junger Herr!</p>

Fig. 11. An example of a translation variation between five editions of Baudissin (1832). Wolff (1926) stands out to be the most distinctive translation among the editions.

to discover patterns or uniqueness in the collection. For example, the domain expert was interested in the word “Lust”. The results are appealing and uncover two main stream of paths that use the word. Most of the translations use the word once (13 translations), four translations did not use this term. As shown in Figure 10, above the top and below bottom main patterns we can see outlier usage of the word. “That is very good because I can see straight away some patterns and most of the translators are using the word in the same segments and some translators are using the word unexpectedly,” the domain expert stated.

Filtering Feedback: The domain expert filters out the outlier translations, and stage direction segments then chooses five different editions of Baudissin (1832). The domain expert discovers that Wolf (1926) [89] and Brunner (1947) [15] show high-Eddy values. The domain expert stated, “These editors ‘intervened’ quite often, altering the text they had received from earlier editions –usually to improve it, i.e. remove bits of poor writing. Both make changes which are not ‘significant’ but just make the text more readable (and actable).” From this focus and similar context, we can depict translation variation easily. For example, in a speech *Iago* says to *Roderigo*: “If thou dost, I shall never love thee after. Why, thou silly gentleman!” All other Baudissin use the word “Freundschaft” which means ‘friendship’ and Wolff (1926) uses the word “Liebe” which means ‘love’ (Figure 11). “Again, intensifying and bringing back the hint of homo-eroticism in the original, which the classic text censored.”

Detailed View Feedback: The domain expert finds this view useful because it fulfills tasks T4, T5. He experimented with this feature and shortly afterwards he started searching for an alignment path he examined previously. He states: “How do I find it now?” The provenance list that stores previously selected paths enables the user to trace back all of the user’s actions whilst exploring the alignment overview. The domain expert finds this helpful to retrace his previous selections. The technical limit to the number of archived user-interactions is the same as that of a web browser with bookmarks. The domain expert might find it cognitively difficult to remember the location of an alignment in the list when there too many archived user-interactions. This cognitive limit can be addressed in future work by enabling the user to personally rename the archived user-interaction labels. Also, the domain expert suggested a potential feature that enables them to keep notes in each alignment. This discussion leads to a new pitfall (PF-33) called feature creep. See Section 7 for details.

Term-Level Comparisons View Feedback: The aligned translations of the base English speech starting with “Thou art sure of me...” show different segments with noticeable higher Eddy values. We easily plot the path of alignments interactively using a drag-and-drop of any segment from the path onto the TLC view. As seen in Figure 3,

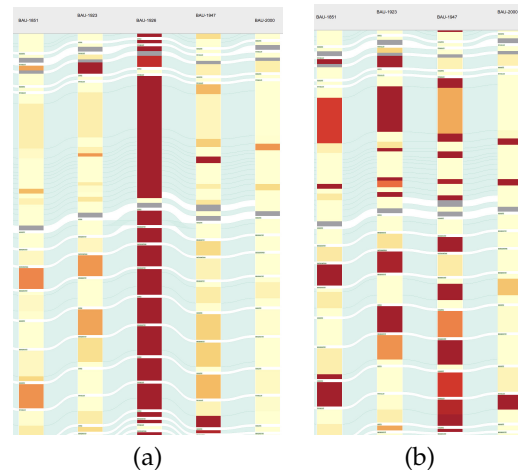


Fig. 12. (a) Five classic canonical Baudissin translations. We can see the Wolff translation stands out due to the high values of Eddy. (b) After de-selecting the 1926 translation, we can see more variation among the four remaining translations.

we discover that there are three translations that stand out from the rest. The three translations have three recognizable words with high values which can be easily observed in the TLC view. These words are translations of the English word “Moor” in the context: “I hate the Moor ...”. The translations are “Maure” in Schwarz (1941) [76], “Schwarze” in Günther (1992) [32], and “Neger” in Buhss (1996) [17]. Most of the translations translate the word into “Mohren”. “That is a good result, that is interesting,” states the domain expert as he looks at the resulting images. He validates the discovery and reasons that as, “That word is translated in different ways in different times with different political implications.”

6.2 Case Studies Using the Design and the Integrated Similarity metrics

We can select a particular set of translations (a sub-corpus) within our corpus on the basis of known features such as type of translation, or date. The Eddy color mapping identifies segments which are of interest because they diverge from others which are more similar. Eddy is calculated on the basis of words, not semantics, so different values do not necessarily predict differences of understanding or interpretation. But Eddy color mapping in this interface encourages an exploratory kind of reading which shifts between scales and between following the course of a single text (vertically) and comparing between texts (horizontally). It also encourages exploring what comparative juxtapositions produce interesting results for a humanistic reader.

Do apparently identical texts differ? Discovering the work of editors/rewriters: As mentioned earlier, our corpus includes different editions of the Baudissin translation, first published in 1832 and often re-published (100s of times so far): printed editions from 1851, 1923, 1926, and 1947, and a digital edition from 2000. Using the interactive drag-and-drop thumbnail view, we can select those five translations. Looking at a distance at this set of five texts, the Eddy value coloring immediately conveys that the 1926 edition is very different from the other four: high Eddy in almost every segment, while the other four show low Eddy values, in nearly all segments as shown in Figure 12(a). This reveals

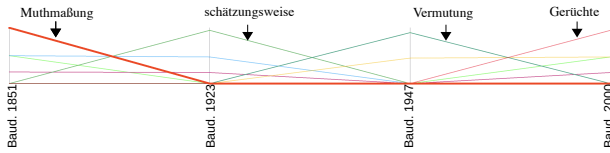


Fig. 13. The TLC view of the path containing the phrase “the aim reports”. The colored lines represent terms and each vertical line represents a translation. The annotated words and arrows illustrate the corresponding terms and lines.

that the 1926 so-called ‘edition’ (‘revised’ by Max Wolff) [89] is not really an ‘edition’ of the Baudissin translation at all: it’s almost a completely new translation, created by editing Baudissin. Interactively zooming in for close inspection shows that Wolff consistently modernizes (rewrites Baudissin in 1920s language, including some slang) and frequently intensifies meanings for dramatic effect, including making the play’s homo-erotic subtext more prominent. This reflects a relaxing of sexual inhibitions in the Weimar Republic.

Next, we can interactively de-select Wolff’s 1926 translation from the sub-corpus under investigation, leaving four actual editions of the Baudissin translation. Now, the updated Eddy coloring shows variation among them –as shown in Figure 12(b), particularly in Brunner (1947) [15] but also others, in the majority of segments. Some of this variation is due to historical spelling changes (lemmatization aims to filter these out, but can still have trouble with some antiquated and poetic, unusual forms of spelling). Some changes are more significant. They are caused by editors aiming to ‘improve’ the Baudissin text they have received, making it easier for readers, actors, and audiences. We can explore variation between these translations at the term level using the TLC view, for example Shakespeare’s phrase “the aim reports” was translated by Baudissin as: “Muthmaßung berichtet” (conjecture reports) (1851, in accord with the 1832 text). This is a very compressed expression (meaning: ‘people making conjectures report’), not easy to follow and also difficult to speak. Some later editors changed this: “schätzungsweise man berichtet” (estimating, people report) (1923), “Vermutung meldet” (supposition reports) (1947), and “Gerüchte melden” (rumours report) (2000). The TLC view (Figure 13) illustrates the translation variation between these four editions. The changes in 1923 and 2000 make the text more like conversational German, less ‘literary’. The change in 1947 substitutes a much commoner word. Such small changes are referred to as ‘silent emendations’ in the history of edition-making. Identifying them is a hideously tedious task for traditional scholarship.

The individual distinctiveness of translators: In the same way, the interface makes it possible to explore comparatively sets of independent translations. Our corpus includes a majority of full-length, poetic translations for theatrical performance –26 of these (including the variant Baudissin texts); also four ‘prose’ translations (for reading, not for performance) and eight theatrical adaptations (shortened, and much freer in the ways they translate). From the interactive thumbnail view, we select the 26, and view them at a distance in chronological order. As seen in Figure 14, the Eddy value colorings clearly show three

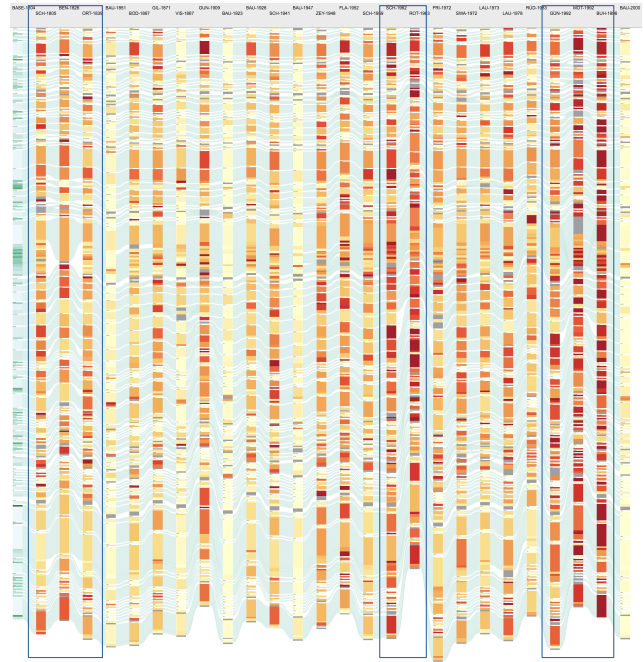


Fig. 14. A sub-set of the translation selected based on the domain expert knowledge shows three periods of generally high Eddy values. The three periods are highlighted by borders.

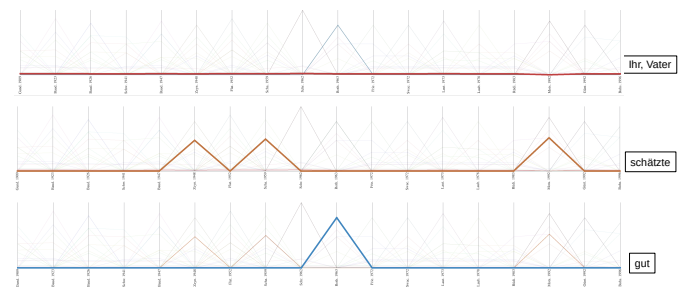


Fig. 15. Three uses of the TLC view to help the domain expert find variation between different translation. The top shows a common use of the words ‘Ihr’ and ‘Vater’. The middle and bottom views show that the words: ‘schätze’ and ‘gut’ which are used only in specific translations.

periods of generally high Eddy value: the early 19th century (Schiller (1805) [73], Benda (1826) [9], Ortslepp (1839) [62]), the 1950s-60s (Schröder (1962) [75], Rothe (1963) [69]) and the 1990s (Günther (1992) [32], Motschach (1992) [59], Buhss (1996) [17]). The translations created before Baudissin (1832) were experimenting with varied ways of translating Shakespeare’s plays. Baudissin’s version of Othello (for the ‘Schlegel-Tieck’ edition of Shakespeare’s plays) soon became canonical - the standard, the one which ‘everybody knows’, even today. Until the 1950s, there were many other German translations of Shakespeare but all were heavily influenced by the ‘Schlegel-Tieck’ style, and Eddy values are low. After the Second World War, new ways of translating Shakespeare began to be tried, as part of a general breaking away from tradition. But this was a complicated process. Innovation and tradition co-existed, tradition becoming stronger again in the 1970s-80s. Then in the 1990s, experimentalism took over.

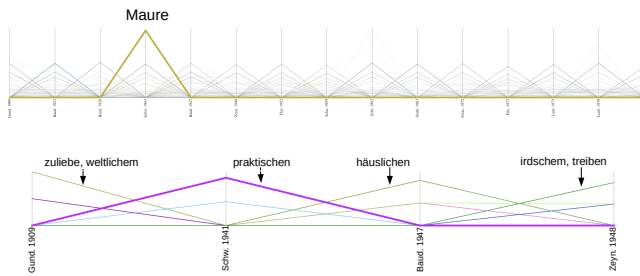


Fig. 16. Two TLC views that show words used distinctively in the translations. In top the word “Maure” was firstly used by Schwarz (1941). In bottom, we can see the distinctive words in the translations of Shakespeare’s odd phrase: “Of love, of worldly matters and direction”.

In the interface, de-selecting the earlier 19th-century translations heightens the visibility of the distinctiveness in the 1950s–1960s and 1990s in terms of Eddy coloring. Inspecting segments with high Eddy in the 1960s and 90s reveals interesting patterns in textual detail. After zooming in for a closer view and with the assistance of TLC view as shown in Figure 15, we can see: Othello’s line beginning ‘Her father loved me’ is translated by all as ‘Ihr Vater liebte mich’ (‘Her father loved me’) except: ‘schätzte mich’ (thought highly of me) (Zeynek (1948) [91], Schaller (1959) [72], Motschach (1992)), ‘war mir gut’ (was fond of me) (Rothe (1963)), ‘mochte mich’ (liked me) (Günther 1992, Buhss (1996)), ‘schien mich zu mögen’ (seemed to like me) (Leonard (2010) [54]). The periods of more different, distinctive translation are evident here. Love and hate, liking and disliking, and their ambiguities, between men and women and between men, are major themes in Othello. Desdemona’s father, Brabantio, hates Othello, his son-in-law. When translators change the kind and intensity of emotions in this way, it matters a great deal. The interface makes it relatively easy to discover patterns in translators’ treatment of emotions and other themes. Exactly why particular translators make particular choices at particular times requires a lot more discussion.

Most of the translators are unknown –their work has never been studied. One is of particular interest because she is the only woman who has translated a number of Shakespeare’s plays into German: Hedwig Schwarz. Her version of Othello is from 1941 [76]. By selecting a subset including other full-length performance versions from her lifetime (1898–1985), we can highlight segments where her version is distinctive. There are not many. In several short segments, she is unusually concise and informal. A feature which stands out is her unusual use of the term ‘Maure’ for ‘Moor’. Othello is a ‘Moor’ –a controversial term, meaning African, variously understood as Black African or North African/Arab/Berber (as in the ‘Moors’ of Moslem Spain). German has two traditional words - ‘Mohr’ meaning Black African and ‘Maure’ meaning North African, with contrary connotations of ‘barbarian’ versus ‘civilized’. All translations before Schwarz used ‘Mohr’. She was the first translator to use ‘Maure’ for ‘Moor’, in the play’s subtitle and in the text. She uses ‘Maure’ when the speaker is respectful, e.g. Othello’s wife, Desdemona. Racist characters (Brabantio, Iago, Rodrigo) use ‘Mohr’ but sometimes ‘Maure’. Rothe

(1963) used the same double-word tactic. No other translators have used ‘Maure’ since. Figure 16(top) shows the TLC view with the word ‘Maure’ highlighted. Some more recent translators use provocatively offensive terms, or neutral terms intended to minimize the racism theme: different responses to ongoing developments in the politicisation of race language and awareness of its emotional force: one of the themes of Othello. Schwarz was a pioneer in using the variety of available German terms in order to dramatize race attitudes within in the play. Her use of ‘Maure’ is clearly intended to dignify Othello as a black man, which is perhaps surprising given when and where she was writing. Interactively reading up, down and diagonally in the corpus enables us to see this.

A further reduced sub-set, using the thumbnail view, can be investigated to compare Schwarz with Baudissin (a translation she knew), Gundolf (1909) [31], one which she may have known, and Zeynek (1947), who also translated during the Nazi dictatorship. Her version is shown to be generally lowest on distinctiveness. She is generally conservative. But Eddy is high for her segment including Shakespeare’s odd phrase ‘Of love, of worldly matters and direction’. We can justify the distinctiveness pattern using the TLC view as shown in Figure 16(bottom). Baudissin had: ‘Der Lieb und unsrem häuslichen Geschäfte’ (love and our household business); Gundolf: ‘Zuliebe weltlichem Geschäft’ (for the sake of worldly business) (a typical poetically condensed translation by him); Zeynek: ‘zu Liebe, irdischem Tun und Treiben’ (for love, earthly activities); Schwarz: ‘für die Liebe, für die praktischen Geschäfte’ (for love, for practical business). The example illustrates her rather practical, easily comprehensible, modern language style. Zeynek, on the other hand, uses antiquated and ornate language, and so scores generally very high Eddy values.

7 STUDY OUTCOME AND DESIGN GUIDELINES

Jänicke *et al.* [47] provide a valuable discussion on collaboration experience with humanities scholars when developing visualization solutions. They report some successful designs and approaches that outline the collaboration process [1]. Jänicke *et al.* present insights and practices that are carried out during the initial project start, the iterative development and the involvement of the digital humanities in it, and the evaluation of the design with the humanities scholars.

This tool has not been fully integrated with the domain expert’s previously used framework due to limited resources. This challenge is reported by Sedlmair *et al.* [77] as pitfall (PF-5: Insufficient time available from potential collaborators). As the domain expert is familiar with visualization projects related to the domain, the project encountered another pitfall (PF-17: Experts focusing on visualization design vs. domain problem) where the domain expert sometimes focuses on the visualization design problems rather than the domain problem. The project also encountered a pitfall (PF-30: Too much domain background in paper) in the earlier stage where we presented unbalanced background towards the domain side. Although we think the domain expert feedback and the case studies are balanced, we observe more emphasis on positive feedback which could indicate to another pitfall (PF-26: Liking necessary but not sufficient for

validation). Table 2 compares and reports our experiences and reflections on the pitfalls reported by Sedlmair et al. Another possible pitfall which is related to the presentation of the work is that attempting to publish the work too soon before serious and thoughtful refinement (PF-32: Premature end: win race vs. practice music for debut).

We also encountered other challenges which are not always a pitfall and might relate to some of the pitfalls reported in Table 2. For example, feature creep [23] (PF-33) is a challenge we encountered, where the domain expert requests features beyond the previously defined requirements and are difficult to incorporate in the current design because it might need another design approach. Another pitfall we encountered (PF-34) is the domain expert expectations of the development cost when implementing new features which may be too high. We believe that this applies to many cross-disciplinary projects where the domain experts lack the knowledge of the development cost.

In order to communicate some of the lessons learned throughout the project lifespan, we derive some general visualization guidelines to facilitate transferring the experience to the general visualization audience:

- Implement support for customization of the texts. Giving the user control of adding or removing translation, and to sort based on a variety of options is always appreciated.
- Provide a mechanism to save the user actions. We think this supports the user better while exploring the application.
- Users appreciate smooth zooming due to the context-and-detail provided in the same window.
- Close reading is always preferable to be available [33], [47]. Also providing users with complete access to the text is beneficial to enable them to analyse and explore text beyond the visual tool.
- Implement a keyword or sentence search functionality. We found this feature useful during the feedback session.
- Support close and distant reading for a large number of texts. We encourage further research that supports distant and close reading for an even larger number of texts.
- Establish good communication with the domain expert(s) in order to guide the project and to provide useful feedback.

We also offer some caution that working on text analysis for multiple languages is very challenging because different languages have different analysis tools associated with them requiring a variety of specialized language-specific knowledge.

8 CONCLUSION AND FUTURE WORK

In this paper, we present a unique, integrated visual design to support distant and close reading of the collection of parallel translations of *Othello*. The visual design aims to present a smooth interactive experience for digital humanities scholars. We identify five main tasks that our proposed application addresses. The application consists of four components. The first one is the context view (alignment overview) of the collection which leverages a range of exploration and interaction techniques. It facilitates smooth zooming which can integrate distant and close reading within the same window. The second component

Pitfall #	Pitfall	How?
PF-5	Insufficient time available from potential collaborators	No time/support for full deployment
PF-17	Experts focusing on visualization design vs. domain problem	Domain expert focuses in design issues
PF-25	Usage scenario not case study: non-real task/data/user	Some reported discoveries made by developers
PF-26	Liking necessary but not sufficient for validation	Domain experts were linked closely during design, however, some criticisms were still reported
PF-30	Too much domain background in paper	At earlier stage, the background was dominant
PF-32	Premature end: win race vs. practice music for debut	Attempts to publish early submissions too soon before serious and thoughtful refinement
PF-33	Feature creep	The domain expert requests features beyond the defined requirements
PF-34	Domain expert high/low expectation	The domain expert might lack the knowledge of the development cost

TABLE 2

A table of the encountered pitfalls identified by Sedlmair et al. [77] and their relevance to this project. The (+) and (-) indicate if we encountered the pitfall during the development of our tool or not.

is the options panel which enables the user to customize the alignment overview. The third is the detailed view which interacts with the main window and displays a close reading of the alignments of the selected segment. Also, it saves the history of user-selected segments and reveals previous alignments. Finally, we present the TLC view, which is a novel and interactive technique to examine the word level variation within the translations. Our proposed application is driven and evaluated by experts from the Arts and Humanities. We provide the domain feedback on the application features. Please see the **supplementary video** for a software demonstration: <https://youtu.be/urMhpDXRL0Q>.

Future work includes generalizability. Our design is applied to a specific corpus and generalizing the platform to adapt different corpus and different languages is problematic because each corpus has different encoding schemes and annotations. Another direction of future work is to incorporate semantic clustering and structural similarity between translations. The rendering could be updated to reflect this when zooming in or out of the scene.

Such a project opens many future possibilities in terms of the scale of the corpus, the use of more sophisticated linguistics or evaluation methods and more languages. We encourage more in-depth research to address the scalability challenge when dealing with a large parallel text dataset. Future possibilities include the application of translation training.

The alignment process is preprocessed and manually generated and is a time-consuming job. Most of the alignment applications use basic and inaccurate algorithms to align two or more segments. Thus, we would like to investigate ways to accelerate the alignment process.

9 ACKNOWLEDGEMENTS

The authors would like to thank Liam McNabb, Dylan Rees and Matthew Roach for their help in proofreading the manuscript. The authors also thank the Technical and Vocational Training Corporation (TVTC) and the Saudi Cultural Bureau for funding and supporting this research endeavour.

REFERENCES

- [1] A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C. R. Johnson, A. Trefethen, and M. Chen. Rule-based visual mappings—with a case study on poetry visualization. In

- Proc. of the 15th Eurographics Conference on Visualization (EuroVis)*, page 381–390, 2013.
- [2] A. Abdul-Rahman, G. Roe, M. Olsen, C. Gladstone, R. Whaling, N. Cronk, R. Morrissey, and M. Chen. Constructive visual analytics for text similarity detection. *Computer Graphics Forum*, 36(1):237–248, 2017.
 - [3] M. Alharbi and R. S. Laramee. SoS TextVis: An Extended Survey of Surveys on Text Visualization. *Computers*, 8(1), 2019.
 - [4] G. Andrienko, N. Andrienko, J. Dykes, S. I. Fabrikant, and M. Wachowicz. Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research. *Information Visualization*, 7(3-4):173–180, 2008.
 - [5] B. Asokarajan, R. Etemadpour, J. Abbas, S. Huskey, and C. Weaver. Visualization of Latin Textual Variants using a Pixel-Based Text Analysis Tool. In *EuroVis Workshop on Visual Analytics (EuroVA)*, 2016.
 - [6] L. Bärfuss. *Othello*. Hartmann & Stauffacher, 2001.
 - [7] W. Baudissin. *Othello, der Mohr von Venedig*. [edited by R Wenig for Project Gutenberg], 1832.
 - [8] M. Behrisch, M. Krstajic, T. Schreck, and D. A. Keim. The News Auditor: Visual Exploration of Clusters of Stories. In K. Matkovic and G. Santucci, editors, *International Workshop on Visual Analytics*, 2012.
 - [9] J. W. O. Benda. *Othello, der Mohr von Venedig*. Shakespear’s dramatische Werke, 1826.
 - [10] Berlin-Brandenburgische Akademie der Wissenschaften. Deutsches textarchiv. <http://www.deutschestextarchiv.de>, 2014. Accessed: 2018-02-10.
 - [11] N. Boyles. Closing in on close reading. *Educational Leadership*, 70(4):36–41, 2012.
 - [12] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2271–2280, 2014.
 - [13] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.
 - [14] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.
 - [15] K. Brunner. *Othello, der Mohr von Venedig*. Englischer Text mit deutscher Übersetzung nach Ludwig Tieck, 1947.
 - [16] M. Büchler, A. Geßner, G. Heyer, and T. Eckart. Detection of citations and textual reuse on ancient greek texts and its applications in the classical studies: eagua project. In *Digital Humanities*, 2010.
 - [17] W. Buhss. *Othello, Venedigs Neger*. Henschel Schauspiel Theaterverlag, 1996.
 - [18] M. Chang and C. Collins. Exploring entities in text with descriptive non-photorealistic rendering. In *IEEE Pacific Visualization Symposium (PacificVis)*, pages 9–16, 2013.
 - [19] T. Cheesman. [delightedbeauty.org](http://www.delightedbeauty.org/). <http://www.delightedbeauty.org/>, 2011. Accessed: 2017-02-16.
 - [20] T. Cheesman, K. Flanagan, and S. Thiel. Translation Array Prototype 1: Project Overview. Technical report, The Arts and Humanities Research Council (AHRC), 2012.
 - [21] T. Cheesman, K. Flanagan, S. Thiel, J. Rybicki, R. S. Laramee, J. Hope, and A. Roos. Multi-retranslation corpora: Visibility, variation, value, and virtue. *Digital Scholarship in the Humanities*, 32(4):739–760, 2017.
 - [22] C. Collins, F. B. Viegas, and M. Wattenberg. Parallel tag clouds to explore and analyze faceted text corpora. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 91–98, 2009.
 - [23] B. Elliott. Anything is possible: Managing feature creep in an innovation rich environment. In *2007 IEEE International Engineering Management Conference*, pages 304–307. IEEE, 2007.
 - [24] E. Engel. *William Shakespeare Othello*. Berlin: Felix Bloch Erben, 1939.
 - [25] R. Flatter. *Othello der Mohr von Venedig. Sonderabdruck für Bühnennzwecke*. Munich: TheaterVerlag Desch, 1952.
 - [26] Z. Geng, T. Cheesman, R. S. Laramee, K. Flanagan, and S. Thiel. Shakervis: Visual analysis of segment variation of german translations of shakespeare’s othello. *Information Visualization*, 14(4):273–288, 2015.
 - [27] Z. Geng, R. S. Laramee, T. Cheesman, A. Ehrmann, and D. M. Berry. Visualizing translation variation: Shakespeare’s Othello. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, S. Wang, K. Kyungnam, B. Benes, K. Moreland, C. Borst, S. DiVerdi, C. Yi-Jen, and J. Ming, editors, *Advances in Visual Computing*, pages 653–663, 2011.
 - [28] M. Gleicher. Considerations for visualizing comparison. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):413–423, 2018.
 - [29] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.
 - [30] V. Gold, C. Rohrdantz, and M. El-Assady. Exploratory Text Analysis using Lexical Episode Plots. In E. Bertini, J. Kennedy, and E. Puppo, editors, *Eurographics Conference on Visualization (EuroVis) - Short Papers*, 2015.
 - [31] F. Gundolf. *Shakespeare in deutscher Sprache*. Berlin: Bondi, 1909.
 - [32] F. Günther. *Othello*. Hartmann & Stauffacher, Ute Nyssen & J. Bausemer, 1992.
 - [33] C. han Jong, P. Rajkumar, B. Siddiquie, T. Clement, C. Plaisant, and B. Shneiderman. Interactive Exploration of Versions across Multiple Documents. In *Digital Humanities*, 2009.
 - [34] M. Harrower and C. A. Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographics Journal*, 40:27–37, 2003.
 - [35] S. Havre, E. Hetzler, K. Perrine, E. Jurrus, and N. Miller. Interactive visualization of multiple query results. In *Proc. of the IEEE Symposium on Information Visualization 2001*, page 105, 2001.
 - [36] J. Heer and G. Robertson. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, 2007.
 - [37] T. Hogan, U. Hinrichs, and E. Hornecker. The elicitation interview technique: Capturing people’s experiences of data representations. *IEEE Transactions on Visualization and Computer Graphics*, 22(12):2579–2593, 2016.
 - [38] S. Howell, M. Kelleher, A. Teehan, and J. Keating. A Digital Humanities Approach to Narrative Voice in The Secret Scripture: Proposing a New Research Method. *Digital Humanities Quarterly*, 8(2), 2014.
 - [39] A. Huang. Similarity measures for text document clustering. In *The sixth new zealand computer science research student conference*, volume 4, pages 49–56, 2008.
 - [40] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *The First IEEE Conference on Visualization: Visualization ’90*, pages 361–378, 1990.
 - [41] S. Jänicke, A. Geßner, M. Büchler, and G. Scheuermann. 5 design rules for visualizing text variant graphs. In *Digital Humanities*, page 12, 2014.
 - [42] S. Jänicke, A. Geßner, G. Franzini, M. Terras, S. Mahony, and G. Scheuermann. TRAViz: A Visualization for Variant Graphs. *Digital Scholarship in the Humanities*, 30(suppl-1), 2015.
 - [43] S. Jänicke, A. Geßner, and G. Scheuermann. A distant reading visualization for variant graphs. In *Digital Humanities*, 2015.
 - [44] J. Johansson and C. Forsell. Evaluation of parallel coordinates: Overview, categorization and guidelines for future research. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):579–588, 2016.
 - [45] B. Jurish. *Finite-state canonicalization techniques for historical German*. PhD thesis, University of Potsdam, 2011.
 - [46] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. In *Eurographics Conference on Visualization (EuroVis) - STARs*, 2015.
 - [47] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. Visual text analysis in digital humanities. *Computer Graphics Forum*, 36(6):226–250, 2017.
 - [48] S. Jänicke, A. Geßner, M. Büchler, and G. Scheuermann. Visualizations for text re-use. In *2014 International Conference on Information Visualization Theory and Applications (IVAPP)*, pages 59–70, 2014.
 - [49] S. Jänicke and D. Joseph Wisley. Visualizing mouvance: Toward a visual analysis of variant medieval text traditions. *Digital Scholarship in the Humanities*, 32:106–123, 2017.
 - [50] D. A. Keim and D. Oelke. Literature fingerprinting: A new method for visual literary analysis. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 115–122, 2007.
 - [51] S. Koch, M. John, M. Wörner, A. Müller, and T. Ertl. Varifocalreader—in-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1723–1732, 2014.
 - [52] T. Korenius, J. Laurikkala, K. Järvelin, and M. Juhola. Stemming and lemmatization in the clustering of finnish text documents. In

- Proc. of the Thirteenth ACM International Conference on Information and Knowledge Management*, pages 625–633, 2004.
- [53] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *IEEE Pacific Visualization Symposium (PacificVis)*, pages 117–121, 2015.
- [54] C. Leonard. *Othello!* Christian Leonard & Shakespeare Company Berlin, 2010.
- [55] L. McNabb and R. S. Laramee. Survey of surveys (SoS)-mapping the landscape of survey papers in information visualization. *Computer Graphics Forum*, 36(3):589–617, 2017.
- [56] H. Mehta, A. Bradley, M. Hancock, and C. Collins. Metatation: Annotation as implicit interaction to bridge close and distant reading. *ACM Trans. Computer-Human Interaction*, 24(5), 2017.
- [57] C. Monroy, R. Kochumman, R. Furuta, and E. Urbina. *Interactive Timeline Viewer (ITLv): A Tool to Visualize Variants among Documents*, pages 39–49. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.
- [58] F. Moretti. *Distant reading*. Verso Books, 2013.
- [59] H. Motschach. *Othello*. Drei Masken Verlag, 1992.
- [60] M. Novotny and H. Hauser. Outlier-preserving focus+context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):893–900, 2006.
- [61] D. Oelke, D. Kokkinakis, and D. A. Keim. Fingerprint matrices: Uncovering the dynamics of social networks in prose literature. *Computer Graphics Forum*, 32(3pt4):371–380, 2013.
- [62] E. Ortlepp. *Othello der Mohr von Venedig*. W. Shakspeare’s dramatische Werke, übersetzt von Ernst Ortlepp, 1839.
- [63] R. L. Ribler and M. Abrams. Using visualization to detect plagiarism in computer science classes. In *Proc. of the IEEE Symposium on Information Visualization*, page 173, 2000.
- [64] P. Riehmman, M. Potthast, B. Stein, and B. Froehlich. Visual assessment of alleged plagiarism cases. *Computer Graphics Forum*, 34(3):61–70, 2015.
- [65] G. G. Robertson and J. D. Mackinlay. The document lens. In *Proc. of the 6th Annual ACM Symposium on User Interface Software and Technology*, page 101–108, 1993.
- [66] S. E. Robertson and K. S. Jones. Relevance weighting of search terms. *Journal of the Association for Information Science and Technology*, 27(3):129–146, 1976.
- [67] A. Roos and T. Cheesman. Version variation visualization (vvv): Case studies on the hebrew haggadah in english. *Journal of Data Mining & Digital Humanities*, 2017.
- [68] R. E. Roth. An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2356–2365, 2013.
- [69] H. Rothe. *Othello, der Maure von Venedig*. Thomas Sessler Verlag, 1963.
- [70] Rui Xu and D. Wunsch. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, 2005.
- [71] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523, 1988.
- [72] R. Schaller. *Othello*. Pegasus Theater- und Medienverlag / Verlag Autorenagentur, 1959.
- [73] F. Schiller and J. H. Voss. *Othello*. Hermann Böhlau Nachfolger / J. B. Metzler, 1805.
- [74] S. Schreibman, A. Kumar, and J. McDonald. The versioning machine. *Literary and Linguistic Computing*, 18:101–107, 2003.
- [75] R. A. Schröder. *Othello*. Suhrkamp, 1962.
- [76] H. Schwarz. *Othello, der Maure von Venedig*. Shakespeare-Bibliothek München, 1941.
- [77] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.
- [78] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proc. IEEE Symposium on Visual Languages*, pages 336–343, 1996.
- [79] H. Siirtola, T. Säily, T. Nevalainen, and K.-J. Rähkä. Text variation explorer: Towards interactive visualization tools for corpus linguistics. *International Journal of Corpus Linguistics*, 19(3):417–429, 2013.
- [80] S. Silvia, R. Etemadpour, J. Abbas, S. Huskey, and C. Weaver. Visualizing variation in classical text with force directed storylines. In *Workshop on Visualization for the Digital Humanities*, 2016.
- [81] E. Suvanaphen and J. C. Roberts. Textual difference visualization of multiple search results utilizing detail in context. In *Proc. Theory and Practice of Computer Graphics, 2004.*, pages 2–8, 2004.
- [82] A. C. Telea. *Data visualization: principles and practice*. AK Peters/CRC Press, 2007.
- [83] M. Wachsmann. *William Shakespeare, Die Tragödie von Othello, dem Mohr von Venedig*. Berlin: Gustav Kiepenheuer Bühnenvertriebs-GmbH, 2005.
- [84] J. Walsh and W. Hooper. Computational discovery and visualization of the underlying semantic structure of complicated historical and literary corpora. In *Digital Humanities*, pages 10–11, 2011.
- [85] F. Wanner, A. Stoffel, D. Jäckle, B. Kwon, A. Weiler, and D. A. Keim. State-of-the-art report of visual analysis for event detection in text data streams. In *Eurographics Conference on Visualization (EuroVis) - STARs*, pages 125–139, 2014.
- [86] D. Wheeler and K. Jensen. Juxta commons. *Digital Humanities*, 5:12, 2013.
- [87] D. R. White and M. S. Joy. Sentence-based natural language plagiarism detection. *Journal on Educational Resources in Computing (JERIC)*, 4:2, 2004.
- [88] C. Wieland. *Othello, der Mohr von Venedig*. Delphine Lettau, 1766.
- [89] M. J. Wolff. *Shakespeares Werke übertragen nach Schlegel-Tieck*. Berlin: Volksverband der Bücherfreunde, Wegweiser-Verlag, 1926.
- [90] J. S. Yi, Y. ah Kang, and J. Stasko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, 2007.
- [91] T. v. Zeynek. *Shakespeare: Othello Der Mohr von Venedig*. Munich: Ahn und Simrock Bühnen und Musikverlag, 1948.



Mohammad Alharbi received his bachelors degree in Computer Science from Qassim University in 2005. He was awarded a Masters in Computer Science from Swansea University specialising in data visualisation in 2011. Currently, he is a PhD candidate at Swansea University and his research focuses on the visualization and analysis of parallel texts.



Robert S. Laramee received a bachelors degree in physics, cum laude, from the University of Massachusetts, Amherst (Zoo Mass) and a masters degree in computer science from the University of New Hampshire, Durham. He was awarded a PhD from the Vienna University of Technology in 2005. From 2001 to 2006 he was a researcher at the VRVis Research Center (www.vrvis.at) and a software engineer at AVL (www.avl.com) in the department of Advanced Simulation Technologies. He was a Lecturer, then Associate Professor at Swansea University. Currently he is an Associate Professor at the University of Nottingham, UK in the School of Computer Science. His research interests are in the areas of scientific visualization, information visualization, and visual analytics.



Zafer Senocak.

Tom Cheesman is Professor of German at Swansea University. Tom is founding Co-Director of the Centre on Digital Arts and Humanities (est.2014). He is Principal Investigator on the collaborative, multi-disciplinary ‘Version Variation Visualisation’ project, a.k.a. TransVis. This project (AHRC-funded) applies Digital Humanities methods to multiple comparable translations. Tom is a founder member of Wales PEN Cymru (2014). His translations from French and German include poetry, fiction and essays by Jörg Bernig, Volker Braun, Esther Dischereit, Ulrike Draesner, Herbert Grönemeyer, and