Characterizing Uncertainty in the Visual Text Analysis Pipeline

Pantea Haghighatkhah, Mennatallah El-Assady, Jean-Daniel Fekete, Narges
Mahyar, Carita Paradis, Vasiliki Simaki, Bettina Speckmann

HAL Id: hal-03784519
https://hal.inria.fr/hal-03784519

Submitted on 23 Sep 2022

Characterizing Uncertainty in the Visual Text Analysis Pipeline

Pantea Haghighatkhah ◇   Mennatallah El-Assady ♣   Jean-Daniel Fekete ♠   Narges Mahyar ‡

Carita Paradis §   Vasiliki Simaki §   Bettina Speckmann ◇

◇ TU Eindhoven ♣ AI Center, ETH Zürich
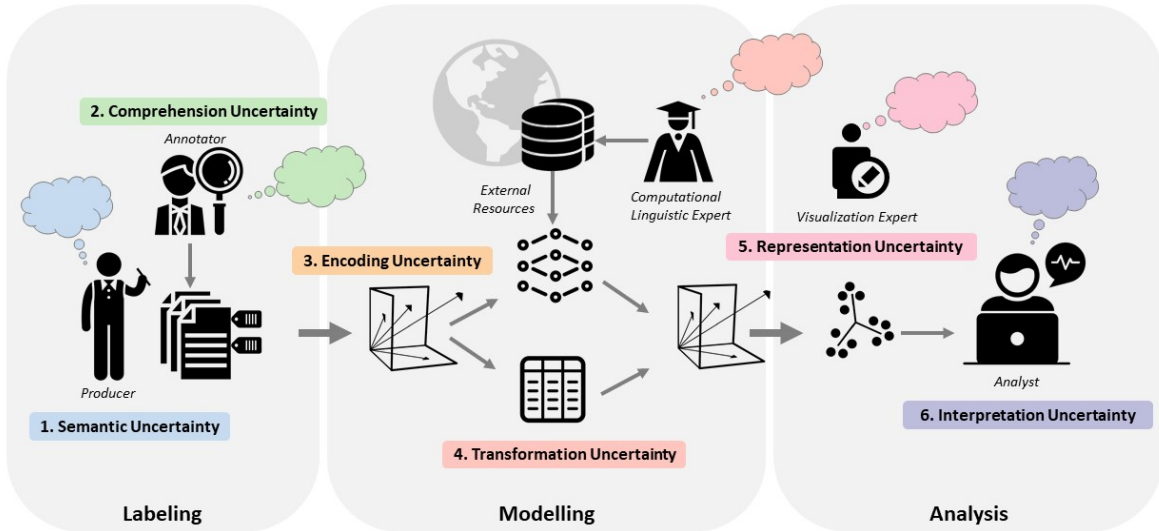♠ Inria & Université Paris-Saclay ‡ University of Massachusetts Lowell § Lund University

Figure 1: We characterize six sources of uncertainty in the three stages of the visual text analysis pipeline.

ABSTRACT

Current visual text analysis approaches rely on sophisticated processing pipelines. Each step of such a pipeline potentially amplifies any uncertainties from the previous step. To ensure the comprehensibility and interoperability of the results, it is of paramount importance to clearly communicate the uncertainty not only of the output but also within the pipeline. In this paper, we characterize the sources of uncertainty along the visual text analysis pipeline. Within its three phases of labeling, modeling, and analysis, we identify six sources, discuss the type of uncertainty they create, and how they propagate. The goal of this paper is to bring the attention of the visualization community to additional types and sources of uncertainty in visual text analysis and to call for careful consideration, highlighting opportunities for future research.

1 INTRODUCTION

With the rapid evolution of natural language processing (NLP) supported by machine learning and other technologies, it becomes possible to analyze and explore large and diverse textual corpora. Visualization is frequently used to support these analyses and explorations. We call the general domain connecting text processing, analysis, and visualization *Visual Text Analysis* [17]. Visual text analysis is used by linguistic experts, domain experts in the social sciences and humanities, or the general public. It is used for various tasks, such as providing an overview, searching, finding related documents, or understanding stylistic patterns within documents, to name a few. All

the analytical and exploratory tasks are done after the text has been processed through multiple steps, as shown in Fig. 1. Each of the steps may, in turn, introduce some level of uncertainty. However, few text visualization systems take into account these multiple uncertainties, in particular, few text visualizations provide cues about them.

Uncertainty may come from different sources, it refers to epistemic situations involving imperfect or unknown information. Uncertainty and error are often used interchangeably. For the purpose of this paper, we consider errors to be deliberate and systemic; they require manual intervention in the processing pipeline to be corrected [3]. If uncertainty is caused by errors in the data, then we generally have no control over it. If the uncertainty is caused by our processing pipeline, then we might or might not be able to precisely identify its source and contribution. In the visual text analysis pipeline, certain processing steps, such as dimensionality reduction, may cause artifacts. These artifacts usually imply that the (visual) representation is not faithful to the underlying data anymore, and as such, they introduce a type of uncertainty. As opposed to errors, artifacts are the result of a deliberate decision by the analyst and can be (automatically) undone by choosing another processing method. Frequently, though, all methods at the analyst's disposal will introduce some sort of artifacts, which can only be mitigated via careful trade-offs.

The main contribution of our paper is a characterization of the multiple sources of uncertainty in the visual text analysis pipeline (see Sect. 3). Not reporting uncertainty generates confusion and causes lack of trust for the users [5]. However, traditional methods for visualizing uncertainty (e.g. [20]) might not always apply to the variety of issues that appear and propagate through the text

analysis pipeline. We hope that our paper will contribute to a better understanding of the uncertainty in the analysis pipeline and motivate future research.

2 RELATED WORK

In this section, we briefly discuss prior work on uncertainty modeling and research on uncertainty visualization. We then describe the importance of modeling and communicating uncertainty in text data.

2.1 Uncertainty Modeling

Uncertainty, error, ambiguity, and other representation issues have been modeled in various ways across research communities. Specifically for linguistic processing and visual analysis, it is imperative to capture such issues for every task at hand. For example, for the task of linguistic annotation, modeling ambiguities and inter-annotator agreement is of utmost importance, as described by Beck et al. [3]. For other tasks, such as sentiment analysis, Bayesian deep learning can be utilized to characterize and model uncertainty [31]. For the task of multi-labeled text classification, Chen et al. [4] quantify uncertainty in the transformation step of the pipeline, deploying experiments to measure both *aleatoric* and *epistemic* uncertainty (various natures of uncertainty). Other approaches focus on uncertainty communication, for example through visualization. For instance, Collins et al. [6] model the uncertainty of text data during visualization using lattice graphs. It's a specific graph-based visualization that expresses the multiple possible outputs that are hidden from the users. While there are many uncertainty modeling approaches, in the context of this paper, we focus on the overall visual text analysis pipeline, including uncertainty visualization, as described in the next subsection.

2.2 Uncertainty Visualization

According to Zuk and Carpendale, uncertainty is a fundamental part of any analytic or reasoning process [32]. Identifying and communicating uncertainty is critical for many analytical tasks. Prior work has addressed the importance of identifying, quantifying, and visualizing uncertainty (e.g., [13, 16]). The complexity of visualizing uncertainty is a known challenge [9]. Researchers in visualization have investigated various techniques for conveying uncertainty through interaction, animation, and sonification [28]. Some examples include Value Suppressing Uncertainty Palettes (VSUPs), which adjust the visual channel allocated to uncertainty based on the level of uncertainty [7], and Hypothetical Outcome Plots (HOPs), that animate a finite set of individual draws [14].

2.3 The Importance of Uncertainty in Text Data

While previous work explores various methods for visualizing uncertainty in general, due to the complexity and ambiguity involved in text data, traditional methods might not always apply to the variety of issues that appear and propagate through the text analysis pipeline. Recent work in text analytics suggests that text visualization needs more careful consideration of uncertainty, its sources, and potential ways to visualize them [12]. The uncertainty in text visualization comes from many origins. First, we need to consider that text is an imperfect representation of human thoughts; therefore, encoding thoughts in a text by nature produces artifacts. Another issue is that people come with different understanding and interpretations of the text. Hence, one single text input can result in multiple interpretations, which affect not only the interpretation of the receiver but also the annotators.

While, Uncertainty can hamper judgment in any decision-making scenarios [29], the impact of uncertainty exacerbates in domains where text data is utilized for important decisions, such as civic decision-making [2], or in humanity research where historians and archivists analyze the text data in newspaper archives to answer fundamental questions about society [10].

The open challenge in the text visual analytics domain is to identify sources of errors and artifacts in the text visualization pipeline and how to design techniques and embed them in various steps of the pipeline to communicate uncertainty to various actors such as annotators and end users (analysts). Due to the inherent complexity of uncertainty visualization, another open challenge is to quantify, model, and visualize the uncertainty to ensure users can understand and interpret it correctly.

3 UNCERTAINTIES IN VISUAL TEXT ANALYTICS PIPELINES

Uncertainty in text visualization stems from many origins. Fig. 1 visually summarizes the six sources of uncertainty we identified in the three stages of the visual text analysis pipeline: labeling, modeling, and analysis. In the following, we describe the different types of uncertainty in detail.

3.1 Semantic Uncertainty

At the start of the pipeline, uncertainty can be introduced by the text *producer*, the way in which the person who types the text or adds a text document to a collection. The errors can be related to the text itself or to the metadata, such as the date of the text production or the text attribution (author). Transcription from a manuscript or from an oral source can always contain transcription errors or misspelled names that will propagate errors and hence uncertainty down the pipeline. Here we specify a type of uncertainty that comes from the producer, and was not caused by textual (e.g misspellings) and metadata errors.

> **Semantic Uncertainty**
>
> The uncertainty caused by the producer's mental linguistic model and translation of thought to text. This uncertainty is also caused by the expression of the speaker's feeling that something is not known or certain, often conveyed by means of specific linguistic markers in the text.

Text is an imperfect representation of human thoughts, therefore encoding thoughts in a text by nature produces artifacts. Fig. 1 shows that the first type of uncertainty concerns the origin of what is communicated through language. This uncertainty refers to the fact that expressions of language are recontextualized thoughts, events, and experiences through language in the written medium. In the transition from the producer through language to the reader, information may or may not be lost. We refer to this as *semantic uncertainty* on the part of the producer. This type of uncertainty cannot always be detected in the text as the producer uses their own linguistic model that can differ from others. There are cases where the semantic uncertainty can be observed from the text with the speaker using language to express "*doubt as to the likelihood or truth of what she or he is saying*" [24], their opinions, facts, or ideas. If producers pay attention to this problem, they often indicate this through expressions such as *may*, *not sure*, *might*, *could*.

Examples: *"I am not sure how to get there.", "We might go to the restaurant.", "We have enough time, haven't we?."* In these example texts, the producers use specific linguistic items (*not sure*, *might*) and ways (tag question *haven't we?*) to express their doubt/uncertainty on the particular way to get to a location, the possibility to go to a restaurant, or whether there is enough time to do something. Please note that if no expressive language for uncertainty is used, then one may not be able to identify the semantic uncertainty of the text.

Challenges: The main challenge at this level is to produce accurate and precise representations in the pipeline of the refined semantic relations expressed in the texts, as well as to identify the uncertainty in the content and how this affects the overall meaning of the text.

3.2 Comprehension Uncertainty

The second source of uncertainty comes from data capturing and annotation performed on the text. This step can include entity recognition and more general annotations such as sentiment analysis, introduced by *annotators*. This enrichment adds higher-level semantics, but also possible errors or misinterpretations that will also potentially change the result of the annotation and introduce uncertainty.

> **Comprehension Uncertainty**
>
> The uncertainty caused during the data collection and annotation process due to technical challenges, perceptual differences and/or other limitations.

People do not respond in the same way to a given text. Hence, one single text input can result in multiple interpretations. This second type of uncertainty that we identified is *comprehension uncertainty*. This uncertainty can be caused during (i) the data capturing and collection process, and (ii) the data annotation process. In (i), the uncertainty is caused by issues related to the representativeness of various text genres/types/contents/topics and their balance in the data set, the noise that irrelevant content creates in the data set, and various biases in the data extraction and collection process. In (ii), multiple factors contribute to the uncertainty, as human annotators come to the task with different linguistic understandings and perceptions of the text. One single text input can result in multiple interpretations, and hence, different decisions by annotators. Phenomena such as polysemy and ambiguity, and vague or generic annotation guidelines can further influence the uncertainty in the annotation process. Also, the annotators' different perceptual systems play an important role in their final decision. All these factors can lead to a high level of disagreement between the annotators, which creates uncertainty about the reliability of the annotated data, and adds more uncertainty in the NLP pipeline.

Examples: Consider a case where the annotator is asked to identify the sense that is addressed or discussed in each sentence e.g. taste, touch, smell, etc. The annotator is given the sentence *"The food is too soft for me."*. It depends on the annotator's comprehension of the sentence to choose *smell*, *touch* (texture) or the *taste* as the associated sense with this sentence. This case concerns the comprehension uncertainty caused by (ii).

Challenges: Apart from the various difficulties that may be faced during the data capturing part, a basic issue that needs to be considered is the design of a solid and well-crafted data collection and annotation protocol, as well as a thorough control (possibly automatic) before finalizing the annotated data set. If this is not adequately addressed, less reliable data is produced, and as a result, more uncertainty is added to the pipeline.

3.3 Encoding Uncertainty

Text encoding in itself can also cause uncertainty. Some complex formats have been designed to encode text in a rich yet faithful fashion, e.g., through the Text Encoding Initiative (TEI) [15]. However, very few projects use these sophisticated mechanisms. More often than not, visual text analysis projects resort to simpler encodings that can lead to information loss or generate ambiguity that causes uncertainty. Even the TEI guidelines, although very rich and well documented, cannot avoid unexpected variations in their interpretation.

> **Encoding Uncertainty**
>
> The uncertainty caused by the data mapping to a data structure, that could lead to a lossy representation of the input.

The TEI guidelines [25], designed to define best practices to encode textual sources with a rich vocabulary of annotations, mention several mechanisms for encoding uncertainty, such as "*levels of certainty*" and "precision" in the chapter "Certainty, Precision, and Responsibility," and encoding for text segments such as "*unclear*", "*gap*" for the transcription of text or speech. All of these textual or linguistic uncertainties are idiosyncratic and intrinsic to our languages, texts, and speech structures. The TEI also allows encoding alternative interpretations for the same text segment (using the $< choice >$ element), as well as marking visible errors ($< sic >$) and possible corrections ($< corr >$). These annotations can become very rich and are currently not supported consistently by visualization systems; they are mostly ignored.

The use of external resources to enrich the text is also a source of errors and uncertainty. For example, when named entities are recognized in a text (e.g., a name or a place name), many NLP systems try to *resolve* them, keeping a dictionary of mentioned persons, looking them up in popular databases (e.g., Wikipedia), or trying to find the location of a named place. These enrichments also lead to errors that consequently lead to uncertainty. For example, the exact address of a person in the 18th century might not be resolved accurately by a modern geocoding service when the street name has changed; the address is then resolved as a city instead of a precise block location.

Example: *One-hot-vector* encoding of the sentence: "*This is an example sentence that contains the word example.*"
- this: $< 1,0,0,0,0,0,0,0,0 >$
- is: $< 0,1,0,0,0,0,0,0,0 >$
- an: $< 0,0,1,0,0,0,0,0,0 >$
- example: $< 0,0,0,2,0,0,0,0,0 >$
- ...

Using a *one-hot-vector* encoding as illustrated above is an efficient way to gather statistical information about the text, but leads to the loss of the word contexts, as the order is not preserved. With such an encoding, uncertainty cannot be reverted, as the original sentence cannot be reconstructed by reverting the encoding.

Challenges: The main challenge for the encoding is to avoid loss of information if this information is usable in the processing. However, some loss is unavoidable since the text data needs to be structured in a form that can be processed. Measuring the information loss during this step can enable an efficient communication of encoding uncertainty. This builds an interesting area for future research.

3.4 Transformation Uncertainty

The encoding typically represents the data as embedding vectors in a high-dimensional space. These get transformed through NLP models that are either exclusively considering the internal data from the pipeline or additionally rely on external resources, such as language modeling or externalizing expert knowledge and feedback. The NLP models introduce transformation uncertainty into the pipeline. Such transformations often rely on design decisions by the *computational linguistics experts*.

> **Transformation Uncertainty**
>
> Uncertainty that is introduced through computations, for example, through language modeling or injection of expert knowledge and feedback.

After encoding and possible enrichment, the text is often transformed to be easy to analyze. Most search engine will transform sentences or documents into high-dimensional vectors using language modeling approaches, such as *word2vec*, *doc2vec*, *BERT*, or *GPT-3*. These vectors allow finding similar documents fast, but they also abstract-out the text and turn it into a representation that humans cannot interpret directly. These transformations are complex

and can generate artifacts, errors, and lead to uncertainty.

Example: "*This sentence is about **cats**.*" and "*This sentence is not about **cats**, but about dogs.*" If we consider the two sentences above and a model that is mapping them to different topics [8], a possible transformation uncertainty can arise if the word *cats* in the second sentence leads to it being partly considered belonging to the topic cats, ignoring the negation. This example is trivial, but more nuanced issues arise in many NLP pipelines.

Challenges: Choosing the appropriate algorithm to model and transform text is crucial for visual text analysis. However, most linguistic models are not fine-tuned to the exact problems they are applied for. Hence, a major challenge is to configure the appropriate processing steps for a given text input and to allow for the capturing of uncertainties in each processing step.

3.5 Representation Uncertainty

The internal representation needs to be presented to a human, and the representation is itself a potential source of uncertainty. For example, visualizing a corpus using a multidimensional computed from the high-dimensional vectors describing the documents is prone to dimensionality reduction artifacts and distortions that can cause uncertainty. These uncertainties are usually introduced by the *visualization experts*.

> **Representation Uncertainty**
>
> In order to visualize the high-dimensional vectors resulting from the transformation step of the pipeline, one needs to use projection techniques to create 2D representation of these vectors. Such process generates artifacts and errors in the distance representation that creates the Representation Uncertainty.

The output of the transformation step is often a high-dimensional vectors that embeds certain information about words, sentences, or documents within itself. Let us assume that in our example pipeline, the transformation step results in vectors that represent documents. The distances between the document vectors capture the relatedness of the corresponding documents. Therefore, a visual observation of the document vectors can be a valuable tool for the purpose of analysis. In order to visualize such vectors, one has to reduce the dimensionality of the vectors to be able to visualize them in 2D. There are various projection techniques that support this such as t-SNE [30] or UMAP [19], however projecting high-dimensional data to lower dimensions is inherently lossy. This implies that the resulting visualization of the low-dimensional projection may be unfaithful to the original distances between the document vectors. Hence, in the 2D representation may carry errors and artifacts that results in the *representation uncertainty*.

In general, when two vectors are close, the documents they represent are similar. However, the 2D representation may include coherent clusters that do not exist in the original data (false neighbors) or conversely miss existing clusters of the original vectors due to topological artifacts resulting from projections (missing neighbors).

A visualization should at least inform the users of these possible artifacts, preferable indicate areas they occur and, if possible, allow resolving them. The lack of communication regarding the representation uncertainty may hinder the trust of the user. Currently, few visualization systems inform users of possible artifacts, and almost none provide techniques to overcome them, especially for many points. Addressing that problem is essential for text visualization in particular, but also for multidimensional visualization in general. Aupetit [1, 11] has proposed a few techniques for small amounts of

data. Martins [18] for larger amounts, but they are not understandable by large audiences.

Example: The artifacts of false neighbors and missing neighbors in the 2D representation can be considered as examples of contributing factors to the representation uncertainty.

Challenges: The most significant challenge is to resolve errors and artifacts generated by this part of the pipeline. However, this is not always achievable due to the lossy nature of the dimensionality reduction methods. Therefore, another interesting task is quantifying the uncertainty and visualizing it in this part of the pipeline. The key here is to come up with an intuitive yet expressive visualization for communication of the errors and artifacts to reduce the representation uncertainty.

3.6 Interpretation Uncertainty

At the last stage of the pipeline the receiver or *analyst* inspecting the visualization is interpreting the data through the lens of the visual design, their interpretation uncertainty is due to the mindset of the *analyst* (user).

> **Interpretation Uncertainty**
>
> The uncertainty caused by the analyst's interpretation of the text visualization at the last stage of the pipeline is called Interpretation Uncertainty.

The representation stage of the pipeline provides the analyst with a visualization of the input text. The analyst has their own linguistic model that they use to interpret the text data. This model is unique to each user (analyst). Besides this, the user has a personal interpretation of the observed visualization that can differ from other humans' interpretations. This can be due to different interpretations of long and short distances and what is scattered and what is a coherent cluster. Similar uncertainties can occur when the analyst is interpreting the colors in the visualization. Both the linguistic model and the visual interpretation of the analyst contribute to the uncertainty generated at the last stage of the pipeline.

Examples *Analyst may interpret a set of points associated with some form of text as a coherent cluster or conversely interpret it as a scattered point set.*

Challenges: Text data is nature is ambiguous, and therefore there is no single canonical data representation or universally accepted interpretation [21, 22]. As a result, the human who reads the visualization and multiple representations of the text can misinterpret the representations. A challenge for the user of the text visualization is to come up with a suitable qualitative or quantitative metric supporting the interpretation of the visualization and therefore reach a more reliable interpretation of the visualization.

Furthermore, the questions that the analyst seeks to answer using the text visualization should be designed in a way that can be answered mostly by quantitative measures and numbers to minimize the effects of personal interpretations.

4 DISCUSSION & CONCLUSION

In this paper, we presented a detailed description of the uncertainty surrounding each step of the visual text analysis pipeline. We did, however, restrict ourselves to the uncertainty of each step in isolation, and the impact of each of the identified uncertainty on future processing steps remains unclear. The steps of the text visualization pipeline can be more intertwined in reality, for instance, the visualization expert and the analyst may revisit the data and output of previous steps of the pipeline. Hence, one of the main challenges

and an opportunity for future work is to characterize the interaction between different steps of the pipeline and the propagation of uncertainty. One plausible assumption is that the propagation *amplifies* previous uncertainties, however, it can also *nullify* errors and artifacts, hiding possible issues that could lead to harmful rationalizations when interpreting the final results [23].

As previous work suggests [4], nature of uncertainty can be characterized as *aleatoric* or *epistemic*. Uncertainty in each step of the pipeline can have either of these natures. In order to tackle it, the first challenge is to identify the nature of uncertainty. If we can identify the sources of uncertainty and their effects throughout the pipeline, the next major challenge is to (visually) communicate their impact. Specifically, exposing all inner workings of the analysis steps and visualizing the errors per step directly, most likely leads to visual and cognitive overload. Possible solutions might be to design in-situ and on-demand explanations through *co-adaptive analytics* [26] by learning from user interactions during analysis and teaching them potential impacts of uncertainty [27].

REFERENCES

[1] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 70(7):1304–1330, 2007. Advances in Computational Intelligence and Learning. doi: 10.1016/j.neucom.2006.11.018

[2] E. P. S. Baumer, M. Jasim, A. Sarvghad, and N. Mahyar. Of Course it's Political! A Critical Inquiry into Underemphasized Dimensions in Civic Text Visualization. *Computer Graphics Forum*, 41(3), 2022. doi: 10.1111/cgf.14518

[3] C. Beck, H. Booth, M. El-Assady, and M. Butt. Representation Problems in Linguistic Annotations: Ambiguity, Variation, Uncertainty, Error and Bias. In *Proceedings of the 14th Linguistic Annotation Workshop*, pp. 60–73. Association for Computational Linguistics (ACL), 2020.

[4] W. Chen, B. Zhang, and M. Lu. Uncertainty quantification for multi-label text classification. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(6):e1384, 2020. doi: 10.1002/widm.1384

[5] J. Chuang, D. Ramage, C. Manning, and J. Heer. Interpretation and Trust: Designing Model-Driven Visualizations for Text Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 443–452. Association for Computing Machinery, 2012. doi: 10.1145/2207676.2207738

[6] C. Collins, S. Carpendale, and G. Penn. Visualization of uncertainty in lattices to support decision-making. In *Proceedings of Eurographics/IEEE VGTC Symposium on Visualization (EuroVis 2007)*, pp. 51–58, 2007.

[7] M. Correll, D. Moritz, and J. Heer. Value-Suppressing Uncertainty Palettes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–11. Association for Computing Machinery (ACM), 2018. doi: 10.1145/3173574.3174216

[8] M. El-Assady, W. Jentner, F. Sperrle, R. Sevastjanova, A. Hautli, M. Butt, and D. Keim. lingvis.io – A Linguistic Visual Analytics Framework. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL): System Demonstrations*, pp. 13–18, 2019. doi: 10.18653/v1/P19-3003

[9] M. Greis, J. Hullman, M. Correll, M. Kay, and O. Shaer. Designing for Uncertainty in HCI: When Does Uncertainty Help? In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, p. 593–600. Association for Computing Machinery (ACM), 2017. doi: 10.1145/3027063.3027091

[10] A. Handler, N. Mahyar, and B. O'Connor. ClioQuery: Interactive Query-Oriented Text Analytics for Comprehensive Investigation of Historical News Archives. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2022. doi: 10.1145/3524025

[11] N. Heulot, M. Aupetit, and J.-D. Fekete. Proxilens: Interactive exploration of high-dimensional data using projections. In *VAMP: EuroVis Workshop on Visual Analytics using Multidimensional Projections*. The Eurographics Association, 2013. doi: 10.2312/PE.VAMP.VAMP2013.011-015

[12] J. M. Hofman, D. G. Goldstein, and J. Hullman. How Visualizing Inferential Uncertainty Can Mislead Readers About Treatment Effects in Scientific Results. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–12. Association for Computing Machinery (ACM), 2020. doi: 10.1145/3313831.3376454

[13] J. Hullman. Why Authors Don't Visualize Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):130–139, 2020. doi: 10.1109/TVCG.2019.2934287

[14] J. Hullman, P. Resnick, and E. Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLOS one*, 10(11):e0142444, 2015. doi: 10.1371/journal.pone.0142444

[15] N. Ide and J. Véronis. *Text encoding initiative: Background and contexts*, vol. 29. Springer Science & Business Media, 1995.

[16] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is My Bus? User-Centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 5092–5103. Association for Computing Machinery (ACM), 2016. doi: 10.1145/2858036.2858558

[17] S. Liu, X. Wang, C. Collins, W. Dou, F. Ouyang, M. El-Assady, L. Jiang, and D. A. Keim. Bridging Text Visualization and Mining: A Task-Driven Survey. *IEEE transactions on visualization and computer graphics*, 25(7):2482–2504, 2018. doi: 10.1109/TVCG.2018.2834341

[18] R. M. Martins, D. B. Coimbra, R. Minghim, and A. Telea. Visual analysis of dimensionality reduction quality for parameterized projections. *Computers & Graphics*, 41:26–42, 2014. doi: 10.1016/j.cag.2014.01.006

[19] L. McInnes, J. Healy, N. Saul, and L. Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018. doi: 10.21105/joss.00861

[20] L. Padilla, M. Kay, and J. Hullman. *Uncertainty Visualization*, chap. 10, pp. 1–18. John Wiley & Sons, Ltd, 2021. doi: 10.1002/9781118445112.stat08296

[21] E. Pavlick and T. Kwiatkowski. Inherent Disagreements in Human Textual Inferences. *Transactions of the Association for Computational Linguistics*, 7:677–694, Nov. 2019. doi: 10.1162/tacl_a_00293

[22] M. Poesio and R. Artstein. The Reliability of Anaphoric Annotation, Reconsidered: Taking Ambiguity into Account. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pp. 76–83. Association for Computational Linguistics (ACL), 2005. doi: 10.3115/1608829.1608840

[23] R. Sevastjanova and M. El-Assady. Beware the rationalization trap! when language model explainability diverges from our mental models of language. In *Communication in Human-AI Interaction Workshop at IJCAI-ECAI'22*, 2022.

[24] V. Simaki, C. Paradis, M. Skeppstedt, M. Sahlgren, K. Kucher, and A. Kerren. Annotating speaker stance in discourse: the Brexit Blog Corpus. *Corpus Linguistics and Linguistic Theory*, 16(2):215–248, 2020. doi: 10.1515/cllt-2016-0060

[25] C. M. Sperberg-McQueen, L. Burnard, et al. *Guidelines for electronic text encoding and interchange*, vol. 1. Text Encoding Initiative Chicago and Oxford, 1994.

[26] F. Sperrle, A. Jeitler, J. Bernard, D. Keim, and M. El-Assady. Co-adaptive visual data analysis and guidance processes. *Computers & Graphics*, 100:93–105, 2021.

[27] F. Sperrle, A. V. Jeitler, J. Bernard, D. A. Keim, and M. El-Assady. Learning and Teaching in Co-Adaptive Guidance for Mixed-Initiative Visual Analytics. In *EuroVis Workshop on Visual Analytics (EuroVA)*, pp. 61–65. Eurographics Association, 2020. doi: 10.2312/eurova.20201088

[28] A. Tse. Why We Are Doing Fewer Interactives. *Malofiej Infographics World Summit*, 2016.

[29] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974. doi: 10.1126/science.185.4157.1124

[30] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

[31] Y. Xiao and W. Y. Wang. Quantifying uncertainties in natural language processing tasks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 7322–7329, 2019.

[32] T. Zuk and S. Carpendale. Visualization of Uncertainty and Reasoning. In *International Symposium on Smart Graphics*, pp. 164–177. Springer, 2007. doi: 10.1007/978-3-540-73214-3_15