# Sentifiers: Interpreting Vague Intent Modifiers in Visual Analysis using Word Co-occurrence and Sentiment Analysis

Vidya Setlur*
Tableau Software

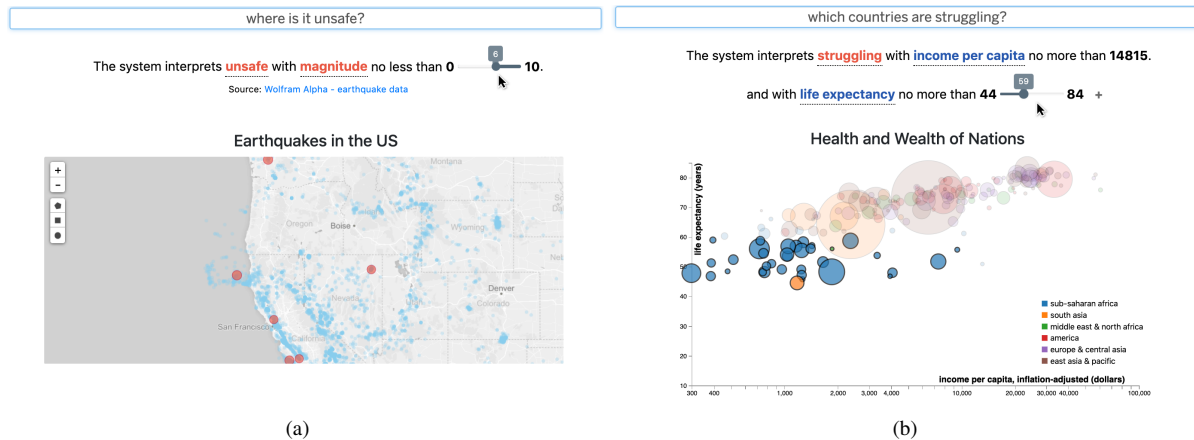Arathi Kumar†
Tableau Software.

Figure 1: Screenshots of *Sentifiers* showing the interpretation of vague intent modifiers using sentiment analysis and word co-occurrence. Interactive text is displayed with the ability to adjust the ranges using slider widgets. (a) For a dataset of earthquakes in the US [38], the system associates the vague modifier "unsafe" with the data attribute `magnitude`. Similar negative sentiment polarities (shown in **red**) result in a `top N` filter of magnitude 6 and higher to be applied. (b) A dataset showing the health and wealth of nations [20]. Here, the modifier "struggling" has a negative sentiment, while the `incomePerCapita` and `lifeExpectancy` attributes have positive sentiments shown in **blue**. The diverging sentiment polarities result in `Bottom N` filters applied.

## ABSTRACT

Natural language interaction with data visualization tools often involves the use of vague subjective modifiers in utterances such as "show me the sectors that are *performing*" and "where is a *good* neighborhood to buy a house?." Interpreting these modifiers is often difficult for these tools because their meanings lack clear semantics and are in part defined by context and personal user preferences. This paper presents a system called *Sentifiers* that makes a first step in better understanding these vague predicates. The algorithm employs word co-occurrence and sentiment analysis to determine which data attributes and filters ranges to associate with the vague predicates. The provenance results from the algorithm are exposed to the user as interactive text that can be repaired and refined. We conduct a qualitative evaluation of the *Sentifiers* that indicates the usefulness of the interface as well as opportunities for better supporting subjective utterances in visual analysis tasks through natural language.

**Keywords:** vague and subjective modifiers, natural language interaction, sentiment analysis, visual analysis.

**Index Terms:** K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

*e-mail: vsetlur@tableau.com
†e-mail: akumar@tableau.com

## 1 INTRODUCTION

Understanding user intent in a query has been recognized as an important aspect of any natural language (NL) interaction system [9, 25]. Search queries typically consist of keywords and terms called *modifiers* that imply a diverse set of search intents [23]. While basic keyword matches from users' search queries might elicit a reasonable set of results, interpreting modifiers provides a better understanding of the semantics in the queries [26].

Recently, NL interfaces for visual analysis tools have garnered interest in supporting expressive ways for users to interact with their data and see results expressed as visualizations [1–3, 15, 19, 22, 30, 35, 36]. Users often employ vague language while formulating natural language queries when exploring data such as "which country has a *high* number of gold medals?" or "what time of the day do *more* bird strikes occur?" [21]. There has been some precedence of research to better understand how these simple vague modifiers comprising of superlatives and numerical graded adjectives should be appropriately interpreted [21, 31]. However, users also employ less concrete and often subjective modifiers such as 'best', 'safe', and 'worse' in utterances [21]. The interpretation of such modifiers makes it challenging for natural language interfaces to precisely determine the extensions of such concepts and mapping intent to the analytical functions provided in the visual analysis systems.

**Contribution**

This paper introduces *Sentifiers*,[1] a system to explore reasonable interpretations and defaults for such subjective vague modifiers in natural language interfaces for visual analysis. The algorithm identifies numerical attributes that can be associated with a modifier using word co-occurrence. Sentiment analysis determines the filter

---

[1]The name *Sentifiers* is a portmanteau of 'sentiment' and 'modifier,' blending their concepts as they co-occur together.

ranges applied to the attributes. Similar polarities result in associating the `Top N` of data values for an attribute with the modifier, while diverging polarities are mapped to the `Bottom N`.

Figure 1a indicates that 'unsafe' and the attribute `magnitude` have similar negative sentiment polarities, defaulting to a higher earthquake magnitude range as seen in the map. The system has the ability to utilize any domain-specific information if available, such as WolframAlpha [4]. Figure 1b shows diverging polarities for the modifier 'struggling' paired with attributes `incomePerCapita` and `lifeExpectancy`. Lower numerical filter ranges based on the statistical properties of the data are applied to generate the scatterplot. Interactive text is displayed to show the provenance of the system's interpretation with clickable portions exposed as widgets that can be refined by the user. An evaluation of the system provides useful insights for future system design of NL input systems for supporting vague concepts in visual analysis.

## 2 RELATED WORK

Research exploring the semantics of vague concepts for understanding intent transcends three main categories: (1) Computational Linguistics, (2) Intent and Modifiers in Search Systems, and (3) Natural Language Interaction for Visual Analysis.

### 2.1 Computational Linguistics

The notion of vagueness in language has been studied in the computational linguistics community [32]. Research has focused on the conceptualization and representation of vague knowledge [8]. The *Vague* system introduces a technique for generating referring expressions that included gradable adjectives [39]. De Melo et al. infer adjective grade ordering from large corpora [14] and Vegnaduzzo automatically detects subjective adjectives [40]. Computational linguists have developed approaches for subjectivity and polarity labeling of word senses [6, 42]. In our work, we draw inspiration from linguistic literature, specifically polarity identification for computing the semantics around vague subjective concepts.

### 2.2 Intent and Modifiers in Search Systems

Search systems have explored techniques to deduce intent in queries during exploratory search. Several techniques exist to extract entity-oriented search intent to improve query suggestions and recommendations [16]. Detecting intent in search systems is also based on query topic classification [33]. Bendersky et al. assign weights to terms in a search query based on concept importance [7]. Recent work has focused on deriving query intent by fitting queries into templates [5, 25]. Li et al. employ semantic and syntactic features to decompose queries into keywords and intent modifiers [25]. Researchers have predicted search intent and intentional task types from search behavior [12, 28]. While the goal of our work to interpret intent in queries is similar to that of search tasks, we focus on resolving vague modifiers to generate relevant visualization responses.

### 2.3 Natural Language Interaction for Visual Analysis

Similar to search systems, natural language interfaces for visual analysis need to understand intent and handle modifiers in the utterances. DataTone provides ambiguity widgets to allow a user to update the system's default interpretation [19]. Eviza and Analyza support simple pragmatics in analytical interaction through contextual inferencing [15, 30]. Evizeon [22] and Orko [35] extend pragmatics in analytical conversation. None of these systems consider how imprecise modifiers can be interpreted. The *Ask Data* system describes the handling of numerical vague concepts such as 'cheap' and 'high' by inferring a range based on the underlying statistical properties of the data [31]. Hearst et al. explore appropriate visualization responses to singular and plural superlatives and numerical graded adjectives based on the shape of the data distributions [21]. We extend this work to more vague, subjective modifiers.

## 3 THE *Sentifiers* SYSTEM

We introduce a system, *Sentifiers* that interprets vague modifiers such as 'safe' and 'struggling' in a NL interface for visual analysis. The system employs a web-based architecture with the input query processed by an ANTLR parser with a context-free grammar, similar to parsers described in [22, 30]. A data manager provides information about the data attributes and executes queries to retrieve data. The query upon execution, generates a D3 visualization result [10].

### 3.1 Algorithm for Interpreting Vague Modifiers

The process for resolving a set of data attributes and their values to a modifier found in the NL input to *Sentifiers*, is outlined as:

---

**Algorithm 1:** Interpretation of Vague Modifiers in *Sentifiers*

---

**Input:** Natural language utterance $\alpha$
**Output:** Generate visualization response
$\alpha$ is the NL input utterance.
$m$ is the vague modifier in the utterance $\alpha$.
Part-of-Speech tagger $POS$ identifies $m$ in $\alpha$.
$attrs_{num}$ is the set of numerical attributes in the dataset $D$.
$attrs_{cnum}$ is the set of co-occurring numerical attributes in $D$
   with $attrs_{cnum} \in attrs_{num}$.
$PMI$ computes co-occurrence scores $w_c$ for $m$ and $attrs_{num}$.
$polarity$ computes sentiment polarities $p$ for $m$ and
   $attrs_{cnum}$.

1  Invoke $POS(\alpha)$ returning $m$.
2  Compute $PMI(m, attrs_{num}) \rightarrow w_c$ for each $attr_i \in attrs_{cnum}$.
3  Compute $polarity(m, attrs_{cnum}) \rightarrow p$.
4  Update interface based on $w_c$ and $p$.

---

#### 3.1.1 Parse Vague Modifiers

Vague modifiers are gradable adjectives that modify nouns and and are associated with an abstract scale ordered by their semantic intensity [24]. Gradable adjectives can be classified into two categories based on their interpretation as measure functions [24]. Numerical graded adjectives such as 'large' and 'cheap' are viewed as measurements that are associated with a numerical quantity for size and cost respectively. Complex graded adjectives like 'good' and 'healthy' tend to be underspecified for the exact feature being measured.

While the interpretation of numerical gradable adjectives has been explored in NL interfaces for visual analysis [21, 30, 31], this paper specifically focuses on the handling of complex gradable adjectives. *Sentifiers* first applies a commonly used performant part-of-speech (POS) tagger during the parsing process to identify these complex gradable adjectives and their referring attributes in the NL utterances [37]. The system can distinguish complex gradable adjectives by checking for the absence of superlative or comparative tags that are used to annotate numerical graded adjectives.

#### 3.1.2 Compute Modifier and Attribute Co-occurrence Scores

The next step maps the vague modifier to a scale based on its semantic intensity so that the modifier can be interpreted as a set of numerical filters for generating a visualization response. We base our approach on linguistic models that represent the subjectivity of complex modifiers as a generalized measure mapping the modifier to numerical attributes in a multidimensional space [18]. For example, the subjectivity of the modifier 'healthy' can be interpreted based on 'weight', 'amount of exercise', and 'hospital visits.'

*Sentifiers* computes the semantic relatedness between the modifier and the numerical data attributes using a co-occurrence measure. To have sufficient coverage for co-occurrence, we use an extensive Google n-grams[2] corpus [27]. To maximize the chances of co-occurrence, *Sentifiers* considers co-occurrence between all n-gram combinations of the modifier and the attribute names. For example,

---
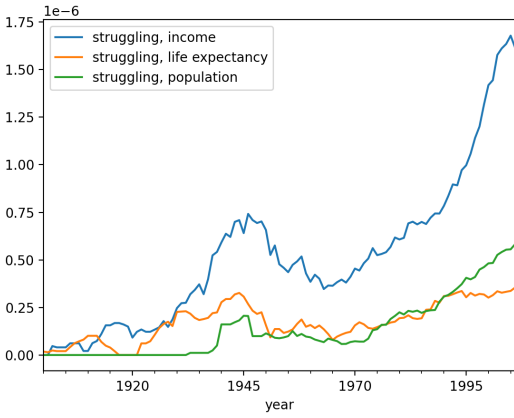[2]An n-gram is a contiguous sequence of n items from a sequence of text.

Figure 2: PMI values for the modifier 'struggling' with each of the attribute n-grams, 'income', 'life expectancy', and 'population' in the Google n-gram corpus. Higher PMI scores indicate a higher co-occurrence of the modifier and attribute terms.

some of the n-grams for the attribute `income per capita` are 'income per capita,' 'income per,' 'per capita,' and 'income.'

We employ a Pointwise Mutual Information Measure (PMI), an information-theoretic measure that quantifies the probability of how tightly occurring a modifier $m$ and a numerical attribute $attr_{num}$ are to the probability of observing the terms independently [13]. We found this measure to work well and was performant with terse word co-occurrence pairings without requiring sentence embeddings. We consider any numerical attribute $attr_{cnum}$ that has a non-zero PMI score, indicating the presence of a co-occurrence with $m$. The PMI of modifier n-gram $t_m$ with one of the attribute n-grams $t_{attr}$ is:

$$PMI(t_m, t_{attr}) = log \frac{p(t_m, t_{attr})}{p(t_m)p(t_{attr})} \quad (1)$$

### 3.1.3 Determine Sentiment Polarities

Once the modifier is semantically associated with co-occurring numerical attributes, we need to determine a reasonable numerical range to associate with the modifier. Sentiment polarity analysis is a linguistic technique that uses positive and negative lexicons to determine the polarity of a phrase [43]. The technique provides the ability to dynamically compute the sentiment of the phrase based on the context in which its terms co-occur rather than pre-tagging the phrase with absolute polarities, which is often not scalable.
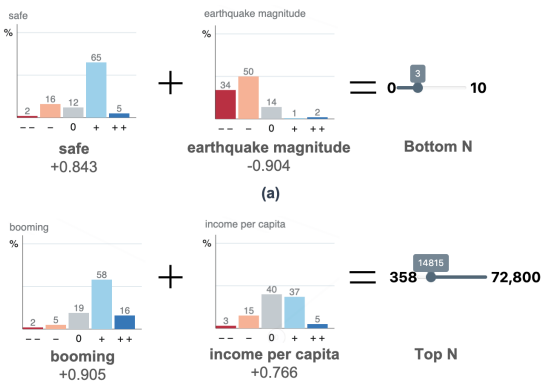


Figure 3: Sentiment polarity logic with sentiments and their normalized scores for the modifiers and the numerical attributes. (a) The modifier 'safe' and attribute `earthquake_magnitude` have positive and negative sentiments respectively, resulting in a `Bottom N` range based on the Richter scale [4]. (b) The modifier 'booming' and attribute `income per capita` both have positive sentiments, resulting in a `Top N` computed based on statistical data properties.

We determine the individual sentiment scores with a sentiment classification based on a recursive neural tensor network [34]. We choose this technique as its models handle negations and reasonably predict sentiments of terser phrases, characteristic of queries to *Sentifiers*. The sentiments are returned as a 5-class classification: very negative, negative, neutral, positive, and very positive. The values are normalized as $[-1, +1]$, ranging from negative to positive to provide an overall sentiment. We then determine the sentiment polarities of the modifier $m$ and co-occurring attribute $attr_{cnum}$ pair based on their individual sentiments (ignoring the strength of the sentiments) using the following combinatorial logic. We treat neutral sentiment similar to positive sentiment as neutral text tends to lie near the positive boundary of a positive-negative binary classifier [43].

**if** ($sentiment_m == positive$ **or** $sentiment_m == neutral$) **and** ($sentiment_{attr_{cnum}} == positive$ **or** $sentiment_{attr_{cnum}} == neutral$) **then**
    Compute $TopN(attr_{cnum})$.
**else if** ($sentiment_m == positive$ **or** $sentiment_m == neutral$) **and** $sentiment_{attr_{cnum}} == negative$ **then**
    Compute $BottomN(attr_{cnum})$.
**else if** $sentiment_m == negative$ **and** ($sentiment_{attr_{cnum}} == positive$ **or** $sentiment_{attr_{cnum}} == neutral$) **then**
    Compute $BottomN(attr_{cnum})$.
**else if** $sentiment_m == negative$ **and** $sentiment_{attr_{cnum}} == negative$ **then**
    Compute $TopN(attr_{cnum})$.
**end if**

*Sentifiers* uses sentiment polarities to compute the ranges in two ways: If domain knowledge exists, the system uses the information to determine a default (Figure 3a uses the Richter scale [4]). Otherwise, the system computes `Top N` to range from $[med + MAD, max]$ and `Bottom N` to range from $[min, abs(med - MAD)]$ where $med$, $MAD$, $min$, and $max$ are the median, median absolute deviation, minimum, and maximum values for $attr_{cnum}$ respectively (see Figure 3b). We choose $MAD$ as it tends to be less affected by non-normality [11].

### 3.2 User Interface

Figure 1 shows the *Sentifiers* interface with an input field that accepts text queries. Upon execution of the query, range filters for the co-occurring numerical attributes are applied, showing a visualization response. The system interpretation is expressed in the form of interactive text [41] above the visualization (Figure 4a) to help the user understand the provenance of how the modifier was interpreted.

Positive, negative, and neutral sentiments are shown in **blue**, **red**, and **yellow** respectively (Figure 4b). The text contains widgets that show ranges starting from the highest co-occurring one. Similar to other NL systems [19, 30, 31], we expose system presumptions as widgets (Figure 4c). If domain-specific semantics are used, a link to the source is provided (Figure 1a). To provide easier readability, *Sentifiers* displays up to two widgets. Word co-occurrence and sentiment analysis techniques can result in incorrect results. The user has the ability repair the system decisions (Figures 4d and f) and the interface updates to reflect the changes (Figure 4e). These refinements are persistent for the duration of the user session.

## 4 EVALUATION

We conducted a user study of *Sentifiers* with the following goals: (1) collect qualitative feedback on the handling of the modifiers for various visual analysis tasks and (2) identify system limitations. This information would help provide insights as to how the handling of complex vague modifiers could integrate into a more comprehensive NL visual analysis interface. The study was exploratory in nature where we observed the types of vague modifiers people asked and how they responded to the system behavior. Because the main goal of our study was to gain qualitative insight in the system behavior, we encouraged participants to think aloud with the experimenter.
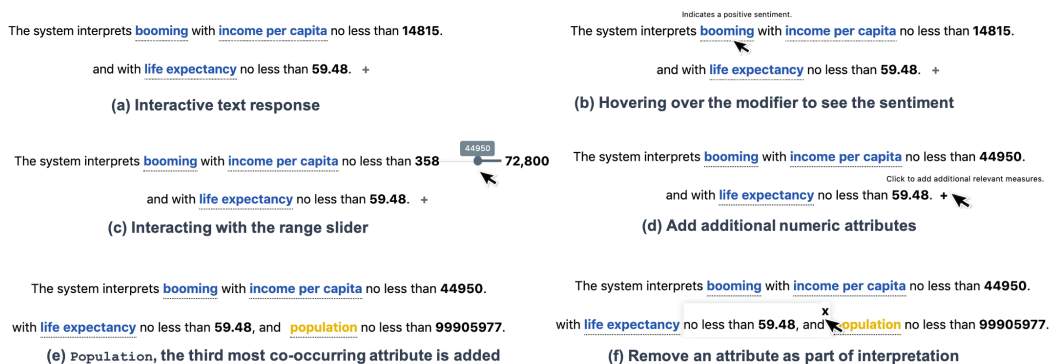
The system interprets **booming** with **income per capita** no less than **14815**.

and with **life expectancy** no less than **59.48**. +

**(a) Interactive text response**

Indicates a positive sentiment.

The system interprets **booming** with **income per capita** no less than **14815**.

and with **life expectancy** no less than **59.48**. +

**(b) Hovering over the modifier to see the sentiment**

The system interprets **booming** with **income per capita** no less than 358 —44950— **72,800**

and with **life expectancy** no less than **59.48**. +

**(c) Interacting with the range slider**

The system interprets **booming** with **income per capita** no less than **44950**.

Click to add additional relevant measures.

and with **life expectancy** no less than **59.48**. +

**(d) Add additional numeric attributes**

The system interprets **booming** with **income per capita** no less than **44950**.

with **life expectancy** no less than **59.48**, and **population** no less than **99905977**.

**(e)** `Population`, **the third most co-occurring attribute is added**

The system interprets **booming** with **income per capita** no less than **44950**.

x

with **life expectancy** no less than **59.48**, and **population** no less than **99905977**.

**(f) Remove an attribute as part of interpretation**

Figure 4: Interactive text response to a query "which countries are *booming*?". *Sentifiers* provides the ability to refine the system defaults.

## 4.1 Method

### 4.1.1 Participants

We recruited ten volunteers (five males, five females, age $24 - 65$). All were fluent in English and all regularly used some type of NL interface such as Google. Eight used a visualization tool on a regular basis and the rest considered themselves beginners.

### 4.1.2 Procedure and Apparatus

Each participant was randomly assigned a dataset of earthquakes in the US [38] or the health and wealth of nations [20] with equal number of participants for each. We began with a short introduction of how to use the system. Participants were instructed to phrase their queries in whatever way that felt most natural and to tell us whenever the system did something unexpected. We discussed reactions to system behavior throughout the session and concluded with an interview. The study trials were done remotely over a shared videoconference to conform with social distancing protocol due to COVID-19. All sessions took approximately 30 minutes.

### 4.1.3 Analysis Approach

We employed a mixed-methods approach involving qualitative and quantitative analysis, but considered the quantitative analysis mainly as a complement to our qualitative findings. The quantitative analysis consisted of the number of times participants used vague subjective modifiers and interacted with the text response.

## 4.2 Results and Discussion

Overall, participants were positive about the system and identified many benefits. Several participants were impressed with the ability of the system to understand their queries ("I typed *scary* to see what it would do, and it understood." [*P*2]). *Sentifiers*' text feedback was found to be helpful ("I wasn't sure how the system would handle this, but it was pretty clear when I saw the response" [*P*4]). The participants appreciated the functionality to be able to correct the system's response ("I wanted to tweak the range a bit and it was useful to be able to change the slider and see the result update" [*P*9]).

The number of unique vague modifiers per participant ranged from 3 to 12 ($\mu = 6.7$) with a total of 24 unique complex modifiers overall. The three most common modifiers were 'good', 'bad', 'severe' for the earthquakes dataset and 'prosperous', 'flourishing', 'poor' for the health and wealth of nations dataset. All participants interacted with the text response to understand the system behavior. The most common interaction was updating the data ranges for the attributes (69% of the interactions), followed by adding new attributes (23%), and deleting attributes from the interpreted result (8%). Comments relevant to this behavior included, "The range seemed high for me and I changed it. It was nice to see the system remember that" [*P*10], "I wanted population to be added to the mix and it was easy to just click and do that" [*P*3], and "I wasn't interested in life expectancy so I just got rid of it" [*P*1].

The study also revealed several shortcomings and provides opportunities for future NL systems supporting visual analysis tasks:
**Support for more complex interpretations:** The current implementation does not support combinations of vague modifiers in the same query. For example, the system was unable to interpret "show me countries that are *doing very well* and *poorly*." [*P*4]. *P*2 expressed that they wanted flexibility in defining analytical functions such as associating 'unsafe' to the frequency of recently occurring earthquakes with magnitude 5 are greater. *Sentifiers* failed to correctly interpret queries such as "which countries are *reasonably doing well*," where *P*7 expected some middle range, though they were able to adjust the ranges after. A comprehensive evaluation with additional datasets would be necessary to ascertain how effective this system would be alongside standard visual analysis tools.
**Handling customization and in-situ curation:** The topic of customization of the interpretation behavior came up during the study. For example, *P*1 said "I typed - show me which countries are *affordable* and it showed me an income range. I was expecting a response that considered inflation, GDP or have a way for me to define that." The algorithms employed in *Sentifiers* assume that the data attributes are curated with human-readable words and phrases. However, data is often messy with domain-specific terminology. Future work should explore mechanisms for users to customize semantics of attributes and interpretations in the flow of their analysis.
**Handling system expectations, biases, and failures:** NL algorithms have shown to exhibit socio-economic biases, including gender and racial assumptions often due to the nature of the training data [17]. Their use can perpetuate and even amplify cultural stereotypes in NL systems. For example, *P*7 commented, "I asked for *good* places to live and the system responded with a high income per capita. To me, that opens up bigger issues such as gentrification and economic segregation." This suggests that there is a responsibility for improved transparency in system behavior; determining appropriate de-biasing methods remains an open research problem.

## 5 Conclusion

This paper presents a technique to explore how a system can interpret subjective modifiers prevalent in natural language queries during visual analysis. Using word co-occurrence and sentiment polarities, we implement *Sentifiers* to map these modifiers to more concrete functions. We expose the provenance of the system's behavior as an interactive text response. An evaluation of the system indicates that participants found the system to be intuitive and appreciated the ability to refine the system choices. Feedback from interacting with *Sentifiers* identifies opportunities for handling vagueness in language in the future design of such natural language tools to support data exploration. As Bertrand Russell stated [29] – "Everything is vague to a degree you do not realize till you have tried to make it precise."

## REFERENCES

[1] IBM Watson Analytics. http://www.ibm.com/analytics/watson-analytics/.

[2] Microsoft Q & A. https://powerbi.microsoft.com/en-us/documentation/powerbi-service-q-and-a/.

[3] ThoughtSpot. http://www.thoughtspot.com/.

[4] WolframAlpha: Computational Intelligence. https://www.wolframalpha.com/.

[5] G. Agarwal, G. Kabra, and K. C.-C. Chang. Towards rich query interpretation: Walking back and forth for mining query templates. In *Proceedings of the 19th International Conference on WWW*, WWW '10, pp. 1–10. ACM, 2010.

[6] C. Akkaya, J. Wiebe, A. Conrad, and R. Mihalcea. Improving the impact of subjectivity word sense disambiguation on contextual opinion analysis. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, CoNLL '11, pp. 87–96. Association for Computational Linguistics, USA, 2011.

[7] M. Bendersky, D. Metzler, and W. B. Croft. Learning concept importance using a weighted dependence model. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pp. 31–40. ACM, 2010.

[8] F. Bobillo and U. Straccia. Fuzzy ontology representation using OWL 2. *International Journal of Approximate Reasoning*, 52(7):1073–1094, 2011. Selected Papers - Uncertain Reasoning Track - FLAIRS 2009.

[9] J. Bos. Computational semantics in discourse: Underspecification, resolution, and inference. *Journal of Logic, Language, and Information*, 13(2):139–157, 2004.

[10] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. In *IEEE Transactions on Visualization and Computer Graphics*, 2011.

[11] P. Cairns. *Doing Better Statistics in Human-Computer Interaction*. Cambridge University Press, 2019. doi: 10.1017/9781108685139

[12] Z. Cheng, B. Gao, and T.-Y. Liu. Actively predicting diverse search intent from user browsing behaviors. In *Proceedings of the 19th International Conference on WWW*, WWW '10, pp. 221–230. Association for Computing Machinery, New York, NY, USA, 2010.

[13] K. W. Church and P. Hanks. Word association norms, mutual information, and lexicography. In *27th Annual Meeting of the Association for Computational Linguistics*, pp. 76–83. ACL, Vancouver, British Columbia, Canada, June 1989. doi: 10.3115/981623.981633

[14] G. de Melo and M. Bansal. Good, great, excellent: Global inference of semantic intensities. *Transactions of the Association for Computational Linguistics*, 1:279–290, 2013. doi: 10.1162/tacl_a_00227

[15] K. Dhamdhere, K. S. McCurley, R. Nahmias, M. Sundararajan, and Q. Yan. Analyza: Exploring data with conversation. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI 2017, pp. 493–504, 2017.

[16] H. Duan and C. Zhai. Mining coordinated intent representation for entity search and recommendation. In *Proceedings of the 24th ACM CIKM '15*, pp. 333–342. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2806416.2806557

[17] B. Friedman and H. Nissenbaum. Bias in computer systems. *ACM Trans. Inf. Syst.*, 14(3):330–347, July 1996.

[18] W. Galit and G. W. Sassoon. Multidimensionality in the grammar of gradability, 03 2016.

[19] T. Gao, M. Dontcheva, E. Adar, Z. Liu, and K. G. Karahalios. DataTone: Managing ambiguity in natural language interfaces for data visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software Technology*, UIST 2015, pp. 489–500. ACM, New York, NY, USA, 2015.

[20] Gapminder. Health and wealth of nations. https://www.gapminder.org/data/, 2020.

[21] M. Hearst, M. Tory, and V. Setlur. Toward interface defaults for vague modifiers in natural language interfaces for visual analysis. In *2019 IEEE Visualization Conference (VIS)*, pp. 21–25, 2019.

[22] E. Hoque, V. Setlur, M. Tory, and I. Dykeman. Applying pragmatics principles for interaction with visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):309–318, 2017.

[23] B. J. Jansen, D. L. Booth, and A. Spink. Determining the user intent of web search engine queries. In *Proceedings of the 16th International Conference on WWW*, WWW '07, pp. 1149–1150. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1242572.1242739

[24] C. Kennedy. *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. Outstanding dissertations in linguistics. Garland, 1999.

[25] X. Li. Understanding the semantic structure of noun phrase queries. In *ACL*. Association for Computational Linguistics, July 2010.

[26] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, USA, 2008.

[27] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, , J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182, 2011. doi: 10.1126/science.1199644

[28] M. Mitsui, C. Shah, and N. J. Belkin. Extracting information seeking intentions for web search sessions. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, pp. 841–844. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2911451.2914746

[29] B. Russell. *The Analysis of Mind*. Library of philosophy. G. Allen & Unwin Limited, 1921.

[30] V. Setlur, S. E. Battersby, M. Tory, R. Gossweiler, and A. X. Chang. Eviza: A natural language interface for visual analysis. In *Proceedings of the 29th Annual Symposium on UIST*, pp. 365–377. ACM, 2016.

[31] V. Setlur, M. Tory, and A. Djalali. Inferencing underspecified natural language utterances in visual analysis. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pp. 40–51. ACM, New York, NY, USA, 2019.

[32] S. Shapiro. *Vagueness in Context*. Oxford University Press, 2006.

[33] D. Shen, J.-T. Sun, Q. Yang, and Z. Chen. Building bridges for web query classification. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pp. 131–138. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1148170.1148196

[34] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642. Association for Computational Linguistics, Oct. 2013.

[35] A. Srinivasan and J. Stasko. Orko: Facilitating multimodal interaction for visual exploration and analysis of networks. *IEEE transactions on visualization and computer graphics*, 24(1):511–521, 2018.

[36] Y. Sun, J. Leigh, A. Johnson, and S. Lee. Articulate: A semi-automated model for translating natural language queries into meaningful visualizations. In *International Symposium on Smart Graphics*, pp. 184–195. Springer, 2010.

[37] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 ACL Conference*, NAACL '03, pp. 173–180. Association for Computational Linguistics, USA, 2003.

[38] USGS. Earthquake facts and statistics. https://www.usgs.gov/natural-hazards/earthquake-hazards/earthquakes, 2020.

[39] K. van Deemter. Generating vague descriptions. In *Proceedings of the First International Conference on Natural Language Generation - Volume 14*, INLG '00, pp. 179–185. Association for Computational Linguistics, 2000. doi: 10.3115/1118253.1118278

[40] S. Vegnaduzzo. Acquisition of subjective adjectives with limited resources. In *In Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, 2004.

[41] B. Victor. Explorable explanations. http://worrydream.com/ExplorableExplanations/, 2011.

[42] J. Wiebe and E. Riloff. Creating subjective and objective sentence classifiers from unannotated texts. In *Proceedings of the 6th CiCLing*, CICLing '05, pp. 486–497. Springer-Verlag, Berlin, 2005.

[43] T. Wilson, J. Wiebe, and P. Hoffmann. Articles: Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3):399–433, 2009. doi: 10.1162/coli.08-012-R1-06-90