



## Review of *Voyant Tools: See through your text*

*Ella Alhudithi, Iowa State University*

---

### Voyant Tools

Sinclair, S., & Rockwell, G.

2021 version 2.4

Web-based text analysis suite

<https://voyant-tools.org>

Free unlimited access

Canada

---



### Introduction

Voyant Tools is an open-source online-based platform for the analysis of digitally recorded texts developed by two humanities computing professors, Stefan Sinclair and Geoffrey Rockwell. Using computational algorithms, the platform extracts linguistic and statistical information from texts of different sizes, types, and languages within seconds. All extractions are available in visual formats (e.g., grids, graphs, and animations) to offer a window for a macroscopic view of texts. This input-output process allows for turning complex metadata into easily interpretable visuals. The platform is freely accessible today, requiring an internet connection and a text collection (i.e., corpus). Users of varying expertise and technical ability can use it to uncover insights that characterize their texts.

### Description

#### Text Entry

Voyant Tools welcomes its visitors with a user-friendly interface that offers four options for text entry (see [Figure 1](#)). The first is to select one of the two preloaded corpora: *Shakespeare* or *Austin*. The *Shakespeare Corpus* consists of 895,737 words sourced from 37 plays written between 1590–1613. The *Austin Corpus* has 781,763 words compiled from eight novels published from 1790–1818. The other options include building a corpus by uploading files, using input boxes, or adding URL links. The latter is the most preferable option for compiling texts automatically from websites. The remaining options are suitable for entering offline stored texts. Many text file types are compatible for uploading into the platform: MS Word, Excel, HTML, RTF, PDF, XML, and Plain Text. To run their analysis, users can upload a single text file, multiple text files, or a zip file with numerous texts. Prior to uploading texts, users should remove any markup tags and irrelevant data, including, but not limited to, page numbers and indices. Completing this clean-up increases the accuracy of automatic word counts. In text analysis, such a stage is critically important due to the extent to which output values can change. Apart from that, the platform accepts texts of any size and type (e.g., a ten-word biography or a million-word newspaper corpus). Likewise, it permits the analysis of texts composed in 13 languages. Clicking on the ‘Language Interface Option’ icon allows for changing the language from English to Arabic, Bosnian, Croatian,

Czech, French, German, Hebrew, Italian, Japanese, Portuguese, Serbian, or Spanish.

**Figure 1**

*Interface of the Text Entry Page*

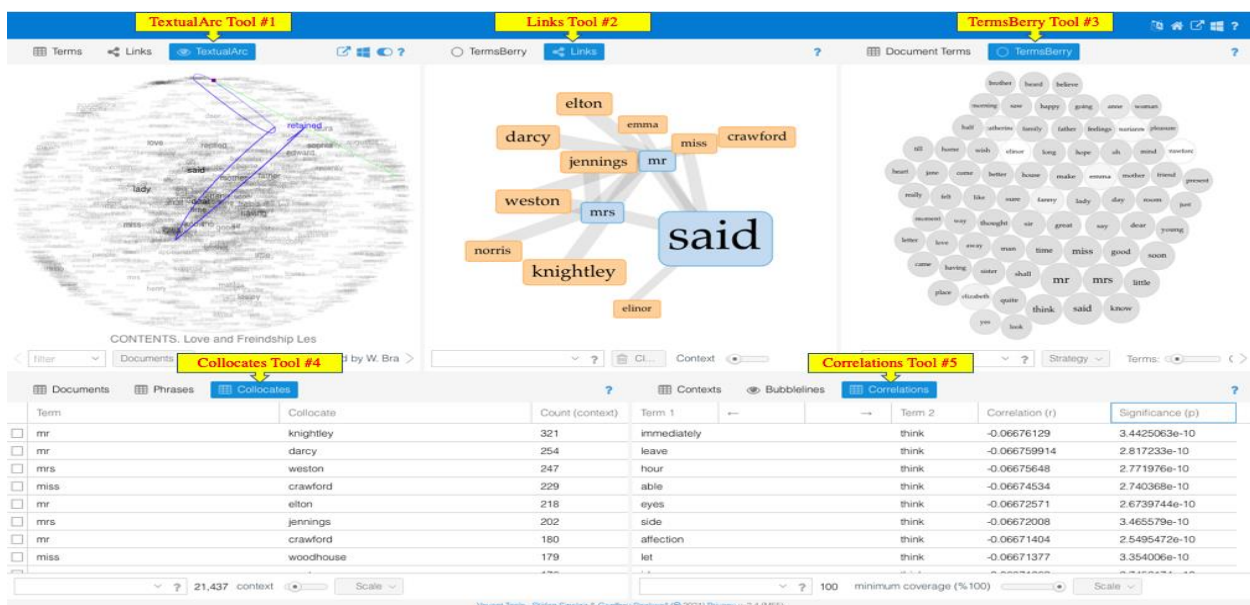
## Text Analytics

Voyant features a package of 29 analytical tools, all supported by highly interactive, rich visualization effects. Clicking on the ‘Tool’ icon from the top right section allows for the selection of different tools. Users interested in commonly co-occurring words can explore the following five tools shown in Figure 2:

1. *TextualArc* offers a dynamic approach to viewing frequent words with others frequently co-occurring together. This tool works by presenting an arrow that moves automatically between words to show their associations.
2. *Links* highlights all connections between high-frequency words through a dynamic network graph.
3. *TermsBerry* introduces keywords with others occurring in proximity, using a sequence of colored bubbles.
4. *Collocates* generates a list of connections that exist between co-occurring words.
5. *Correlations* reveals significant positive and negative correlations among frequent words, using Pearson’s and regression coefficients.

**Figure 2**

*Panels of Analytical Tools Targeting Words Commonly Used Together*

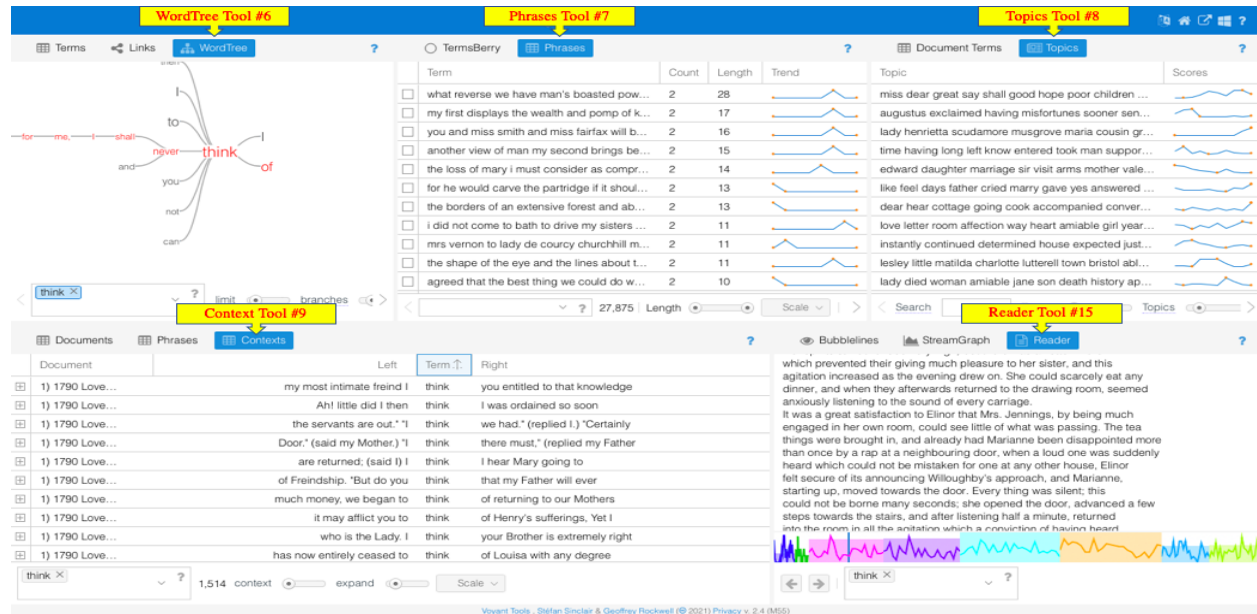


The platform also offers ten tools for qualitative explorations of contexts (see [Figure 3](#)).

6. *WordTree* offers a sophisticated way of viewing phrases in which words appear.
7. *Phrases* is a tool that outlines all common word sequences, with a length ranging between 2–30 words per sequence.
8. *Topics* automatically identifies connections among content words.
9. *Contexts* presents concordance lines showing utterances surrounding frequent words.
10. *Mandala* features a network graph that links every word to all texts that it appears in.
11. *Terms* reports all words that are unique to a particular text.
12. *Microsearch* prints out blocks of different sizes to represent texts and to show their lengths.
13. *Dreamscape* is a map that displays all geographical places mentioned in texts.
14. *RezoViz* is a network graph highlighting all connections found among people, places, and organizations.
15. *Reader* provides a window for reading all texts with a line graph showing word frequencies and distributions.

**Figure 3**

### Tools for Context-Focused Analysis



Most of the remaining tools are available for a more inclusive view of word counts (see [Figure 4](#)).

16. *Document* provides estimates of word types, tokens, type-token ratios, and means.
17. *Summary* analyzes all texts based on word frequency, length, density, and distinctiveness.
18. *DocumentTerms* shows word distributional patterns with raw and relative frequencies.
19. *CorpusTerms* offers details regarding words occurring in a corpus, like collocates, correlations, and phrases.
20. *Knots* gives a more creative way of identifying common words by displaying lines that twist depending on occurrence counts.
21. *TermsRadio* presents a line graph that depicts changes in word counts.
22. *StreamGraph* is a visualization that highlights significant changes in word counts.
23. *Trends* divides texts into ten equal segments to demonstrate patterns of word use.

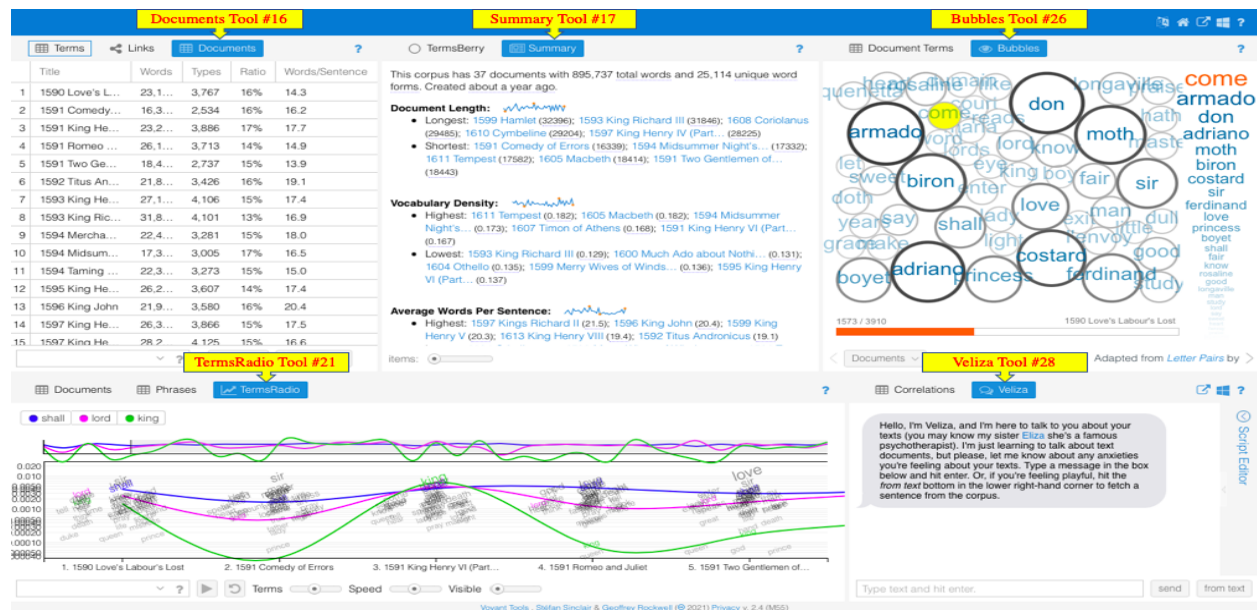
24. *Cirrus* presents a cloud that displays the top 25–500 words where higher frequency words have larger fonts in the center.
25. *Bubblelines* compares different levels of word frequency by displaying animated bubbles of varying sizes and colors.
26. *Bubbles* presents all words according to their order in texts using a sequence of moving bubbles.
27. *ScatterPlot* incorporates several statistical analyses into one to create different word clusters.

The last two tools provide assistance in setting up results and running analyses.

28. *Veliza* is a digital guide that generates replies based on user's written inquiries.
29. *Catalogue* functions as a database interface for filtering search criteria.

## Figure 4

Tool Panels Presenting Numerical Values and Assistance Guides

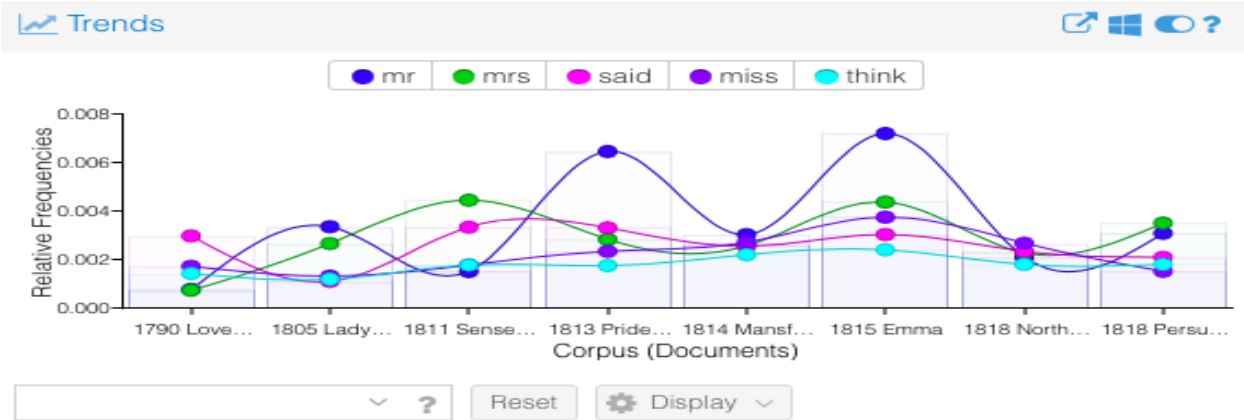


## Functionalities

Voyant Tools permits users to customize various features of its interface and analytics. By default, the analysis page features five panels that represent different tools: *Cirrus*, *Reader*, *Trends*, *Summary*, and *Context*. The top section of every panel presents four option icons: 'Help', 'Options', 'Tool', and 'Export', whereas the bottom part offers a search box and a scale display (see Figure 5). When pressing any icon or display, a drop-down menu with available functions appears. All these functions are adjustable based on user preference. To better trace patterns, users can minimize the scale to a single text segment from the bottom section. For a more inclusive view, users have different options for enlarging the scope. One way to expand it is by typing word roots followed by an asterisk in the search box. Other strategies include filtering the scale display by changing the number of analyzed texts and words. On-screen help is available in the top-right section, offering detailed step-by-step instructions, screencast tutorials, and core design principles for all tools. The remaining portion of the panel features several visual effects, ranging from basic grids to multidimensional dynamic graphs. From the 'Options' icon, users can customize their sizes and their colors to better highlight features of interest. As an illustration, red symbols can represent higher frequency counts, whereas green ones can indicate lower frequencies.

**Figure 5**

*A Tool Panel Featuring Four Option Icons, a Search Box, and a Scale Display*



Another function of Voyant Tools is the creation of categories to classify words. Clicking on the 'Options' icon allows for forming and editing word lists. Among the many categories, users can classify content words based on their positive or negative semantic meanings (see Figure 6). Other lists can target functional words (e.g., pronouns, articles, and prepositions). Having such categories enables users to hide or display their patterns of use. An equally important function is the selection of word frequency units: raw (i.e., total) or relative (i.e., normalized). While the first is useful for recording frequencies in a single text, the latter is preferable for counting frequencies in two or more texts due to the adjusted weighted values. Another primary function of the platform is the ability to export outputs. Such an option allows for retrieving information from any completed analysis without re-uploading texts to access lost outputs. Pressing the 'Export' icon presents users with all available options. For visualized outputs, users have the chance of exporting photos in three forms: PNG, SVG, and URL. The latter works best for tools featuring dynamic, animated visualizations. Exporting alphanumeric outputs is available in several forms: PDF, TXT, PNG, SVG, and URL. Finally, users have the opportunity to apply any customized function globally to all tools or to one tool.

**Figure 6**

*A Drop-Down Sub-List for Word Categories*

The figure shows a screenshot of the "Categories" interface in Voyant Tools. It features two columns: "positive" and "negative". Each column has a list of words and a plus (+) or minus (-) button. At the bottom, there are buttons for "Add Category", "Remove Selected Terms", "Cancel", and "Save".

Category	Word
positive	freedom
	advantage
	excellent
	superior
	confidence
	enjoy
	wonderful
negative	bad
	concern
	fail
	despair
	desperate
	disadvantage
	depression

## Evaluation

### Technicality

Voyant Tools is entirely web-based and does not require any login or installation. Given its open access, the most important function is the flexibility to perform any text analysis task using any device (smartphone, tablet, or computer) and operating system (Android, Mac OS, Linux, iOS, or Windows). Another aspect that increases the value of the platform is the transparent design with a simple input-output process that users can navigate with ease. The interface of all featured tools is consistent, with the same customizable options and operating functions. Although expected, one difficulty facing users not versed in the domain can be the terminology. The system employs many technical terms, such as pipelines, collocates, and density. Such technicality might hinder the process of navigation and discovery. For users looking to grow their competence, the 'Help' icon offers enriching opportunities for exploring a range of relevant research applications, teaching demonstrations, and screencast tutorials.

### Discovery

Voyant Tools offers a unique environment for text analysis, with automated computational approaches and multidimensional visualized reports. When entering texts, the platform produces varying textual descriptions within seconds, including, but not limited to, single words, collocates, clusters, and phrases. The platform also permits the exploration of statistical information (e.g., word ratios, distributions, and significances). With recent calls for a more transparent analysis approach that supports language specialists, researchers, teachers, and students (Breyer, 2009; Chambers, 2019; Jewell & Zillig, 2011; Römer; 2010), the platform achieves this by automatically performing all computations. That is, users have the advantage of selecting any feature of interest from the option icons without setting up parameters to extract information. Such operations make the tools indeed efficient for a wide range of users and analysis purposes. Since its release, the platform has been employed in course instructions (Buzarna-Tihenea, 2020; Joubert, 2021), assessment reports (Hendrigan, 2019; Miller, 2018), and research projects (Hetenyi et al., 2019; Ming, 2018). Another useful and innovative aspect that facilitates the analysis is the various visual effects available for highlighting patterns and connections. Offering users different colors, shapes, and symbols increases the possibility of identifying trends and retrieving meaningful conclusions. Such visualizations can surely promote the discovery of salient features characterizing texts. Another notable benefit is the chance for users to rearrange results to meet their analysis goals. For users interested in historical-linguistic shifts, for instance, filtering search criteria to prioritize relative frequency counts and segmentations can highlight meaningful longitudinal patterns. Likewise, users analyzing linguistic variations in a corpus of published research articles might benefit from sorting results from the most occurring three-word sequence to the least frequent ones. While such results may be helpful, users should rearrange them with caution to avoid selective reporting and subjective interpretation.

### Practicality

Among the many affordances of Voyant Tools, one worth noting is the recognition of textual information recorded in up to 13 languages. This capability can provide users with enriching insights about text complexity within and across languages. The opportunity to navigate both the interface and the tools using different languages increases the practicality while minimizing any linguistic difficulty encountered. Apart from that, the platform does not require sophisticated knowledge in the setting of search parameters, making it usable by any user, including novices. However, those eager to invest time in learning such principles can take advantage of the 'Help' icon. Another useful property of the platform is that it allows for exporting both textual and visual outputs without the need to copy and paste or type entries into files. This feature reduces the human effort and time constraint while saving the alignments of generated outputs. One shortcoming is that users cannot perform any modifications after exporting outputs and closing web pages. Users, therefore, have to ensure that their scales are well adjusted, values are correctly weighted, and entries are clearly highlighted prior to exiting any page. Another noteworthy limitation related to the usability of Voyant Tools is the protection of data. With the absence of



documentation on storing information in the system, analyzing sensitive data (e.g., consent statements and health reports) is likely to be high-risk. Therefore, users analyzing such data have to be cautious as personally identifiable information might be improperly recorded. The last notable drawback of the platform is the absence of published word lists, such as Academic Word List (Coxhead, 2000), New General Service List (Browne, 2013), and Academic Vocabulary List (Gardner & Davies, 2014). Embedding such lists would generate meaningful comparisons across different texts and accommodate a range of research purposes. Ultimately, the developers acknowledge the possibility that shortcomings might emerge, resulting in publishing all scripts in GitHub to open opportunities for collaboration and improvement.

## Summary

Voyant Tools is an online-based environment for frequency-based analysis of computer-readable texts. The environment is enriched with 29 visualization tools that retrieve linguistic and statistical features within seconds. This instant production of numerous features makes the platform potentially useful for diverse research and teaching applications. With the rich affordances and extensive documentation, users of various experience levels and technical abilities can efficiently use the platform to transform complex metadata into comprehensive visuals that characterize their texts. This is especially true in cases involving texts with millions of words. Overall, the easy-to-use design, the diverse analytical suite, and the data-driven approach provide valuable opportunities for users to move beyond close readings of texts to discover patterns in a more quantifiable and less subjective manner.

## References

- Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153–172. <https://doi.org/10.1080/09588220902778328>
- Browne, C. (2013). The new general service list: Celebrating 60 years of vocabulary learning. *The Language Teacher*, 4(37), 13–16. <https://doi.org/10.37546/JALTTLT37.4>
- Buzarna-Tihenea, A. (2020). Text analysis tools in ESP teaching. *Economic Sciences Series*, 2, 252–258. <https://doaj.org/article/4716f2bc1aeb45f99cc7c8235b5a325b>
- Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, 52(4), 460–475. <https://doi.org/10.1017/S0261444819000089>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>
- Hendrihan, H. (2019). Mixing digital humanities and applied science librarianship: Using Voyant Tools to reveal word patterns in faculty research. *Issues in Science and Technology Librarianship*, 91. <https://doi.org/10.29173/istl3>
- Hetenyi, G., Lengyel, A., & Szilasi, M. (2019). Quantitative analysis of qualitative data: Using Voyant Tools to investigate the sales-marketing interface. *Journal of Industrial Engineering and Management*, 12(3), 393–404. <https://doi.org/10.3926/jiem.2929>
- Jewell, A., & Zillig, B. (2011). Counted out at last: Text analysis on the Willa Cather Archive. In A. Earhart & A. Jewell (Eds.), *The American literature scholar in the digital age* (pp. 169–205). University of Michigan Press. <https://doi.org/10.3998/etlc.9362034.0001.001>
- Joubert, E. (2021). Distant reading in French music criticism. *Nineteenth-Century Music Review*, 1–25. <https://doi.org/10.1017/S1479409820000476>

- Miller, A. (2018). Text mining digital humanities projects: Assessing content analysis capabilities of Voyant Tools. *Journal of Web Librarianship*, 12(3), 169–197. <https://doi.org/10.1080/19322909.2018.1479673>
- Ming, X. (2018). Smart learning models of certified legal translators and interpreters in China. *Comparative Legilinguistics*, 36(1), 47–64. <https://doi.org/10.14746/cl.2018.36.3>
- Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present, and future. In M. C. Campoy-Cubillo, B. B. Fortuño, & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 18–38). Continuum. <https://doi.org/10.1093/elt/ccq080>

### About the Author

Ella Alhudithi is a doctoral student in Applied Linguistics and Technology at Iowa State University, where she also instructs academic writing in the ESL program. Her research interests are in language pedagogy, corpus-based discourse analysis, and languages for specific purposes.

**E-mail:** [ella@iastate.edu](mailto:ella@iastate.edu)