

Researchers, Teachers, and Learners Seeing New Possibilities with Voyant Tools

August 2020. Volume 24, Number 2

Title	Voyant Tools, v. 2.4 (M29)
Developers	Stéfan Sinclair & Geoffrey Rockwell
Website	https://voyant-tools.org/
Product	A text mining tool
Hardware Requirements	An internet-accessible device (computer, tablet, smartphone)
Operating Systems	Cross-platform
Online Help	https://voyant-tools.org/docs/
Price	Free

Text mining is a process where users interact with a corpus to extract information from unstructured textual data by identifying and exploring patterns dependent on preprocessing procedures, various algorithms, and visualization tools (Feldman & Sanger, 2007). The result of this discovery process is systematic structuration of text via automatized searching, indexing, clustering, and classification operations (Ananiadou et al, 2010).

These days, text mining is increasingly popular in English language research and education (Warschauer, Yim, Lee, & Zheng, 2019). Here, I introduce Voyant Tools as a text mining software used for reading and analyzing texts. In the hope of sparking curiosity and satisfying diverse interests, after presenting a description and evaluation of the software, this article explores possible applications for researchers, teachers, and learners in the field.

Description

Overview

As described by the developers, Voyant Tools (v. 2.4) is a free “web-based reading and analysis environment for digital texts” (Sinclair & Rockwell, 2020). It is an open-source

collection of 20+ interactive tools, which allows for data extraction and provides textual and statistical analysis to uncover insightful patterns or trends within or across texts. Below, I describe inputs, outputs, types of tools, exports, and where to find help.

Inputs

As a text mining software, Voyant Tools is dependent on textual input. It can instantly analyze text of different genres and modes (i.e., written texts or transcripts of spoken communications) on diverse topics in documents of various formats (plain text, HTML, XML, PDF, RTF, and MS Word and Excel) even in archives such as .zip files (see Voyant Tools Help for further details, Figure 1).



Figure 1. Voyant Tools Help Page.

Preloaded collections are available for experimentation. However, for a more meaningful and relevant experience, we can add our own corpus of customized texts directly (by inserting them or pasting URLs) or upload files from our computers to the input page (Figure 2).

We may choose to set a few options (e.g., the language, the process for determining unit boundaries, and the stopwords list) before revealing a dashboard display of the corpus. This can be done using the Language Interface Options icon or Options toggle at the top right of the input box.

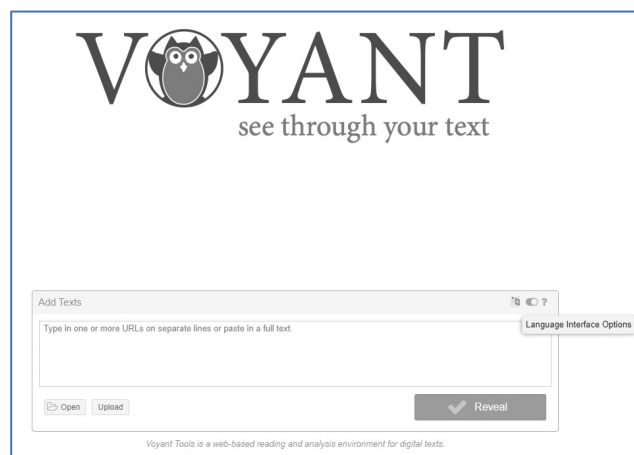


Figure 2. Voyant Tools Input Page.

Outputs

The dashboard display, or ‘skin’ (Figure 3), has five default panels featuring the Cirrus, Reader, Trends, Summary, and Contexts tools. These offer a selection of textual, tabular, and visual representations (charts, graphs, and networks) of the data. The following example shows unpublished learner data from a small, university academic writing class aimed at developing paragraph-writing skills.

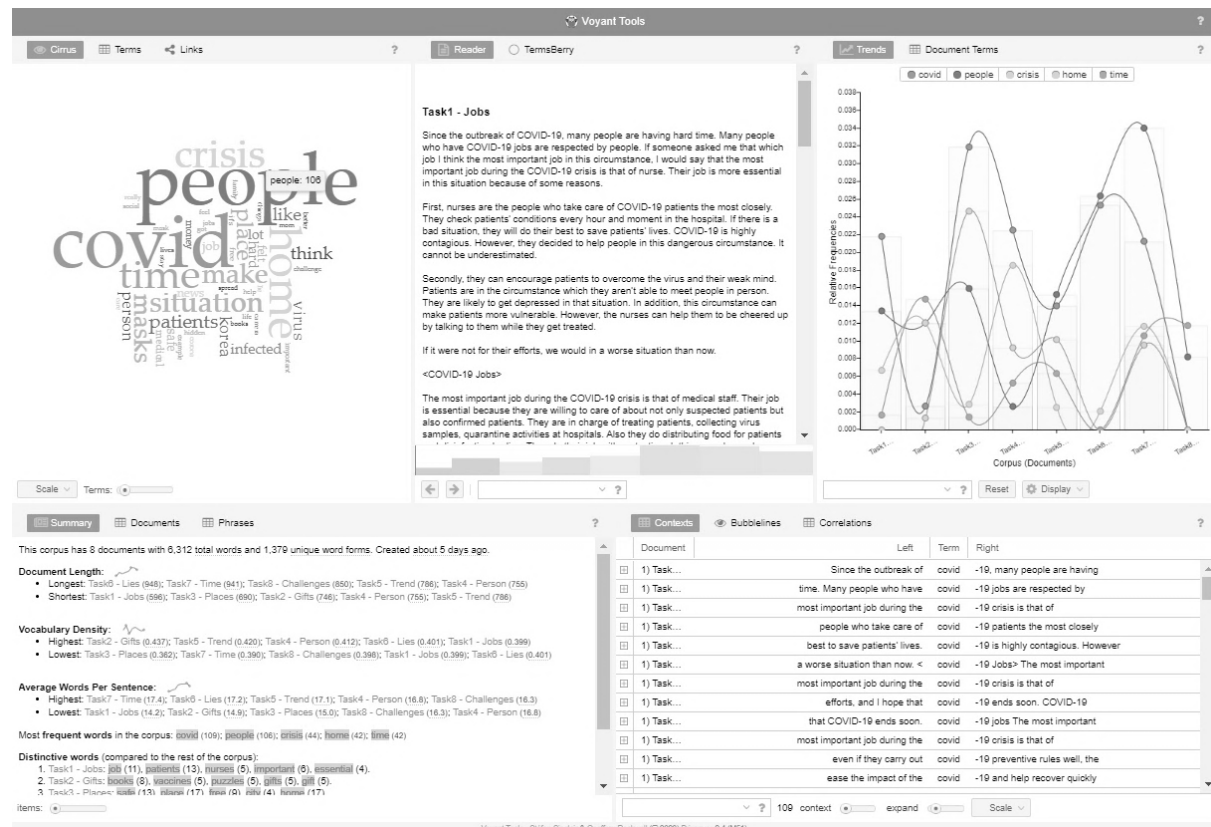


Figure 3. Voyant Tools Dashboard Display.

Collection of Tools

The panels can be customized via the Window icons, which allow users to access further tools (Figures 4 and 5) to study terms, collocates, contexts, correlations, relationships, distributions, and frequencies or repetition in texts. The various tools often have additional adjustable search and filtering settings (see toggles and dropdown lists below the tools).

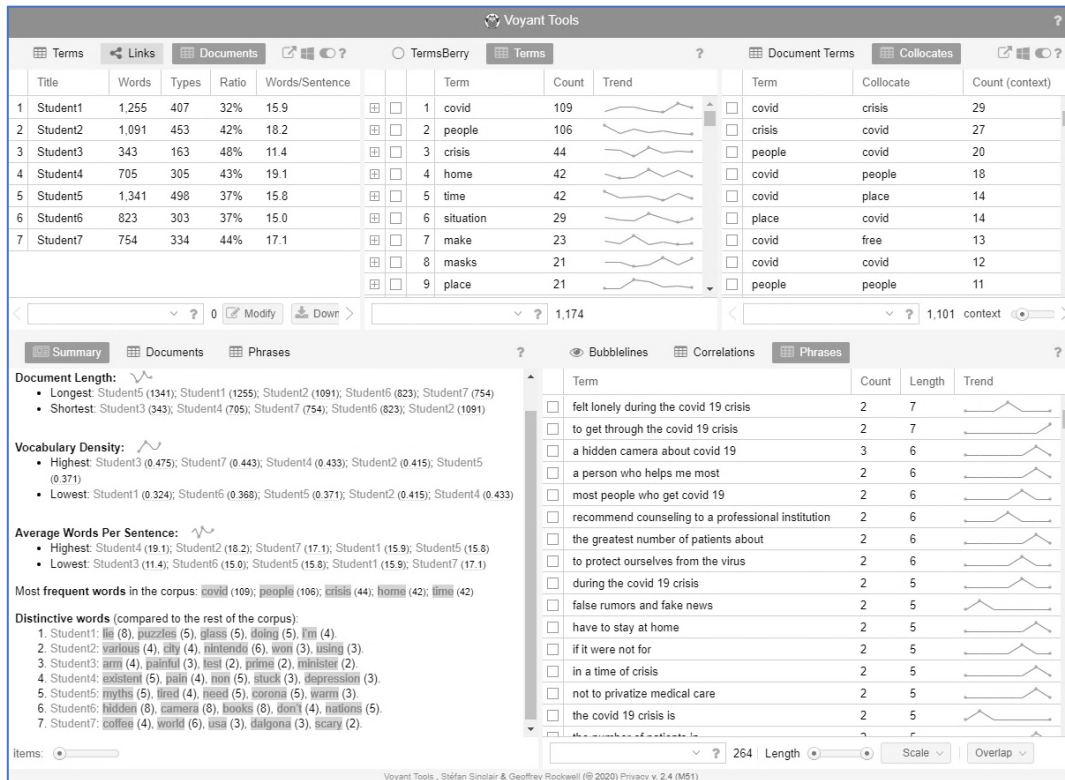


Figure 4. Textual Tools (Collocates, Documents, Summary, Phrases, Terms).

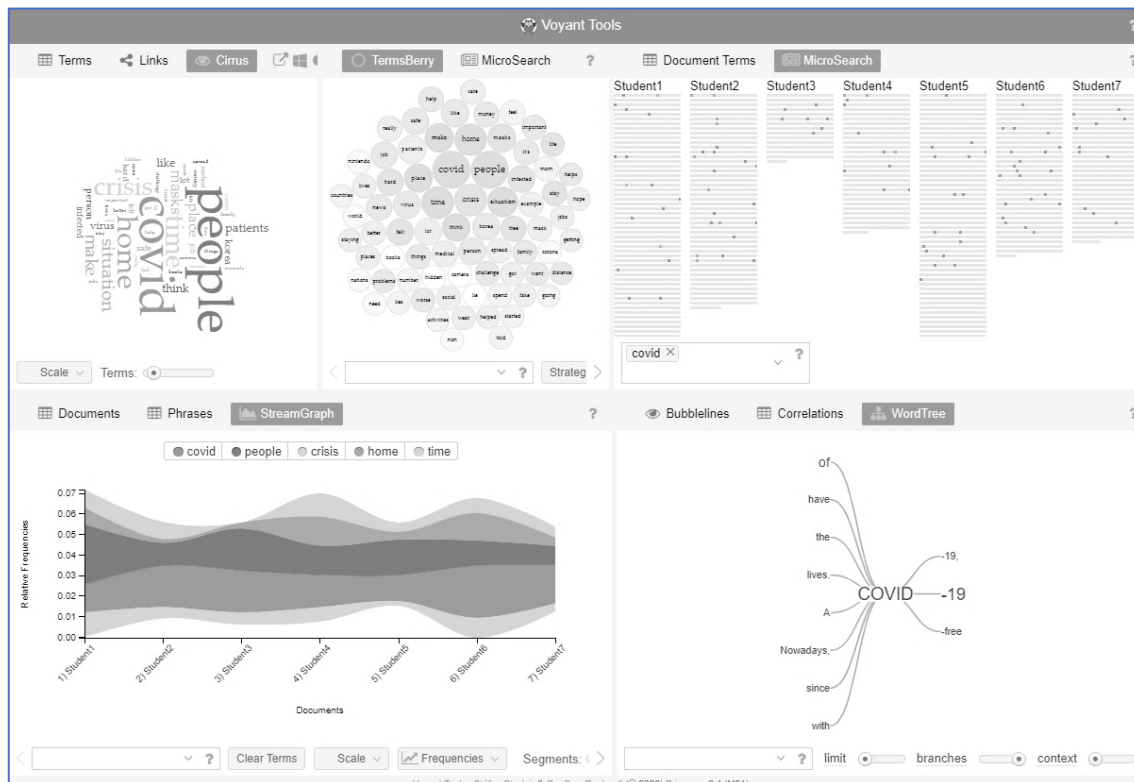


Figure 5. Visualization Tools (Cirrus, MicroSearch, StreamGraph, TermsBerry, WordTree).

Exports

Depending on the type of display, the content of each panel can be exported in image (PNG or SVG) or tabular form (via URLs or HTML) by pressing the Export Arrow icon above each tool. The same arrows also allow outputs to be embedded into other websites.

Help

Clicking the Question Mark icons above each tool enables users to access Voyant Tools Help, which additionally features detailed descriptions of most tools (Figure 6).

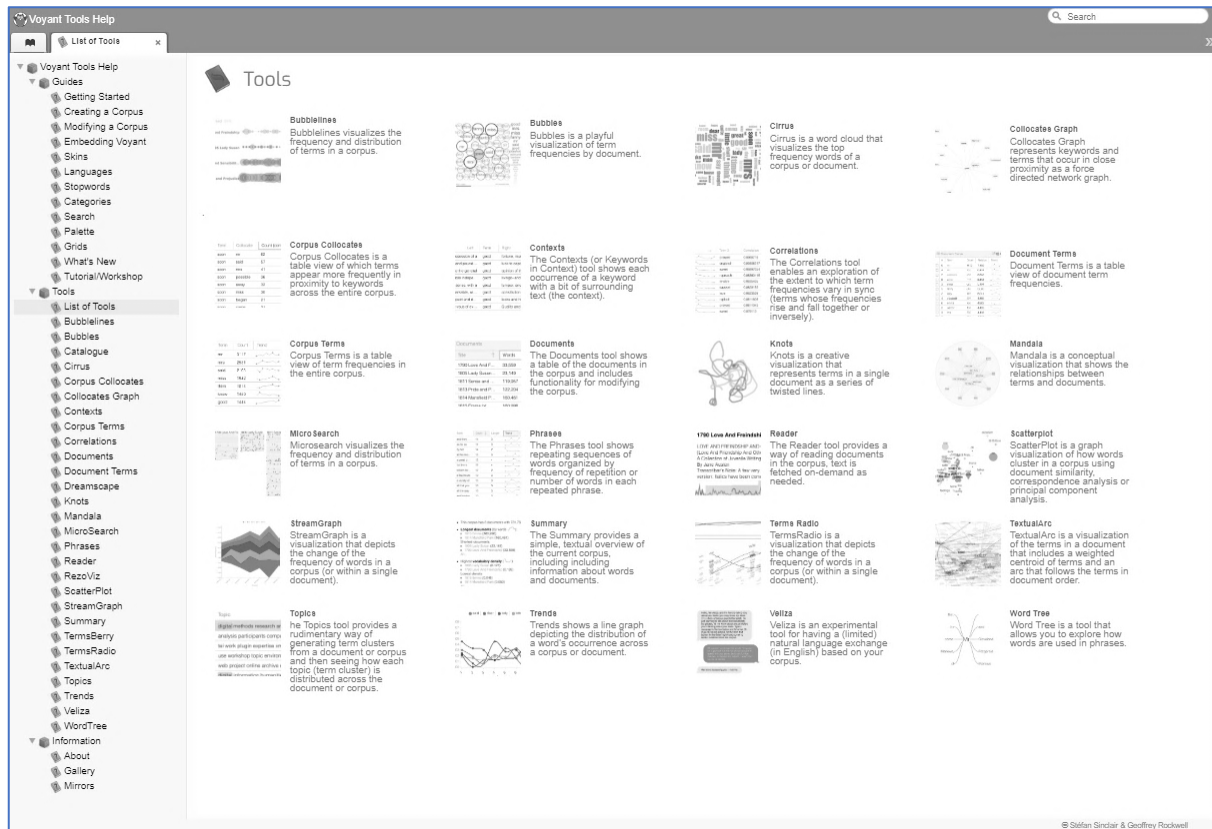


Figure 6. Voyant Tools List of Tools.

Evaluation

Having given a general description above, I would like to make comments regarding the users, the interface and tools, the presentation, data selection and preparation, and limitations of analysis.

Users

As stated in the About section of Voyant Tools Help, the software was designed not only for academics but also for students in the digital humanities and the general public (Sinclair & Rockwell, 2020). This implies an adult audience. However, use of Voyant Tools can be adapted to younger learners and non-academic needs.

Interface and Tools

This rapid text mining software has a simple input-output interface. Nevertheless, navigating certain tools may not be intuitive to all. For example, those in Figures 7 and 8 may be

especially challenging to read and interpret. Thus, practice and discernment are needed to skillfully select tools and customize displays for meaningful analysis.

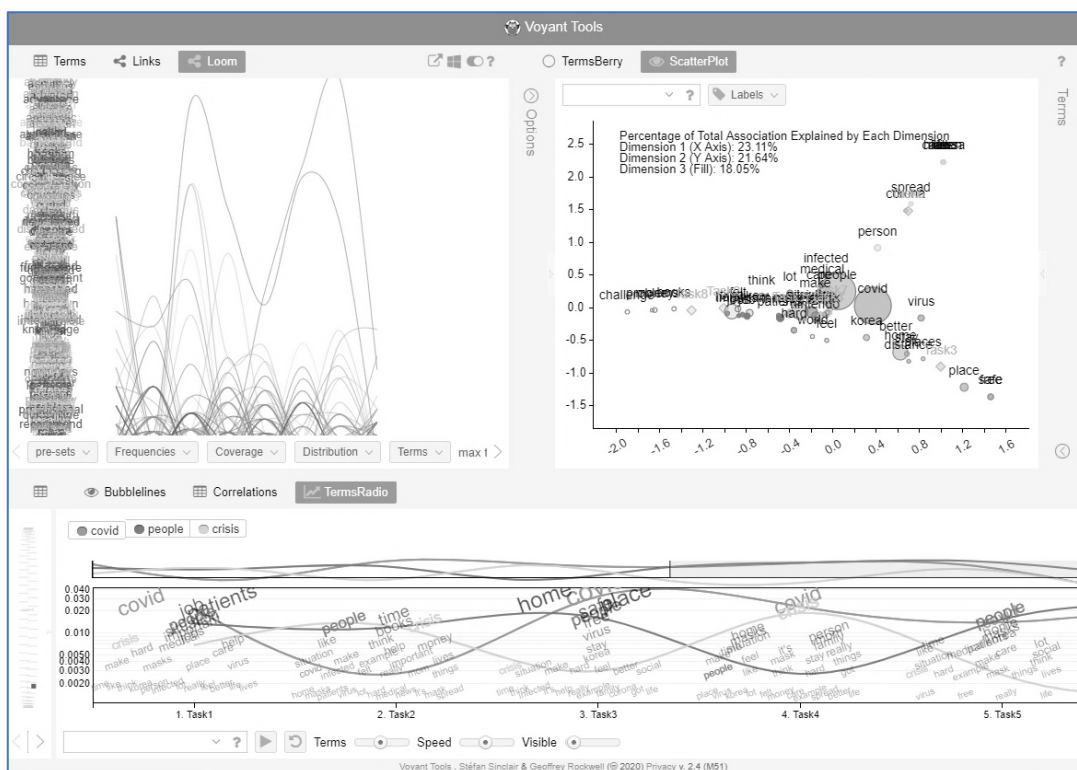


Figure 7. Detailed Visualization Tools (Loom, ScatterPlot, TermsRadio).

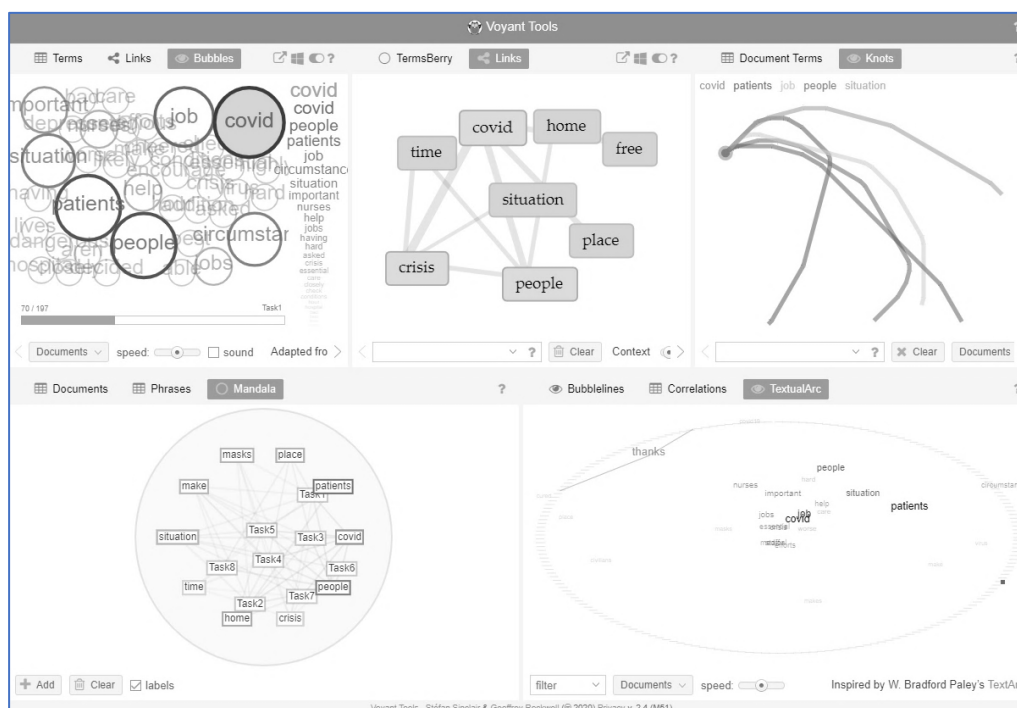


Figure 8. Movement Tools (Bubbles, Links, Knots, Mandala, TextualArc).

Presentation

Voyant Tools is more comfortably viewed on bigger screen devices than smaller ones. Yet, numerous didactic possibilities exist for synchronous or asynchronous activities. If learners have access to computers, they can themselves interact with the interface. Otherwise, teachers can display outputs in presentations. Another option is to export specific outputs to a mobile-friendly application or to provide printed materials.

Data Selection and Preparation

When it comes to language education contexts, data usually consists of a ‘pedagogic corpus’ (Willis, 2003, p. 163) of authentic or constructed textual classroom materials. However, studying user-generated data could also be interesting. Depending on the research or teaching/learning purpose, inputs have to be selected carefully.

Another consideration is corpus size. Sinclair and Rockwell (2020) only vaguely indicate the number of documents (one or many) that Voyant Tools can analyze. A general guideline is that if a corpus is too small, results may not be very informative or reliable, although sometimes “smaller, more specialised and context-specific corpora” (Walsh, 2011, pp. 99) produce more meaningful results. In the case of Voyant Tools, the choice of tools seems to be decisive as “Some of the functionality depends on multiple documents and some tools work less well when there are hundreds or more” (Sinclair & Rockwell, 2020).

Before running the software, texts may need some clean-up. For instance, word and sentence boundaries should be clearly identified. Also, working with texts which contain alternate spellings, mistakes, contractions, abbreviations, acronyms, or unusual punctuation, if these are not important, may require some manual standardization.

Limitations of Analysis

While analyses may be interesting and revealing, one limitation the designers acknowledge in the Languages section of Voyant Tools Help is that the tools have “very little language-specific functionality, such as part-of-speech tagging” (Sinclair & Rockwell, 2020). Additionally, the software cannot group words by family. A further limitation is that some tools are still experimental. For example, those in Figure 9 need further refining.

Applications

One area of application for Voyant Tools is data-driven learning (DDL). DDL, which originated in the field of corpus linguistics in the 1990s, is essentially a hands-on pedagogical approach to developing language awareness (Gilquin & Granger, 2010) that relies on text mining. It employs corpus tools and techniques, especially concordancers, to reveal how language is used by getting learners to notice frequent linguistic patterns as well as combinations of grammar and lexis (Pérez-Paredes et al., 2019). Given the versatility of Voyant Tools, applications for the software abound for research, lesson planning and materials development, and classroom learning related to DDL.

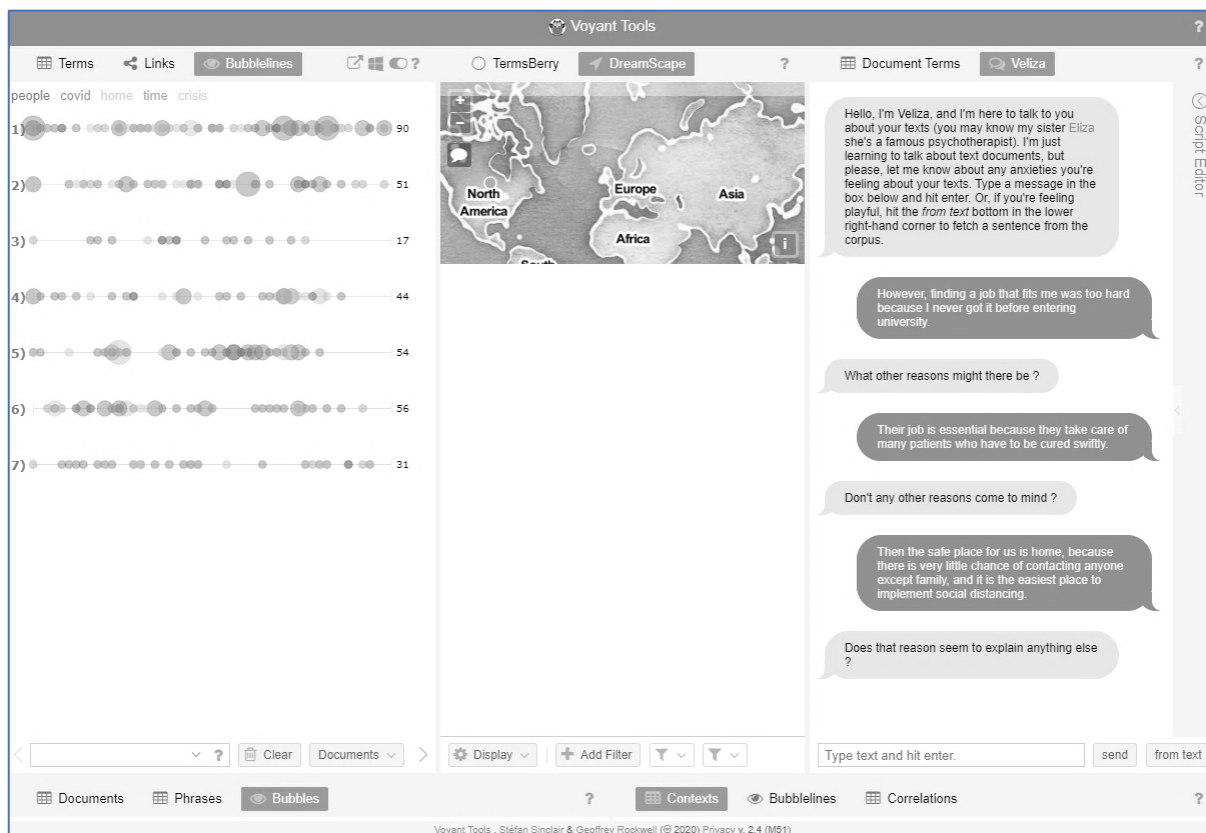


Figure 9. Experimental Tools (Bubblelines, Dreamscape, Veliza).

Research

Voyant Tools is generally useful for research (and teaching research methods) in applied linguistics, and there are multiple ways to use the software. For instance, in the context of action research projects on teacher instructions, feedback, or student-teacher discourse, it could be used for deductive purposes (for verification of hypotheses) or inductive ones (to discover fresh insights). To explore learner production, the tools could be used to study cross-sectional group tendencies (e.g., similarities and differences in vocabulary choice, key word use patterns over the course of similar texts, or word use in context) in performing specific tasks or assignments. Alternatively, the software could be used to observe longitudinal trends in individual output (e.g., preferred terms, sentence length, and vocabulary density). Because the software enables technical analyses of large quantities of texts, it is particularly convenient for distant readings when the goal is lexical or lexicogrammatical analysis. However, it can additionally be used as a complement to close readings of texts as in mixed studies. By informing teaching, research using Voyant Tools could be valuable for teacher training purposes or for professional development.

Lessons and Materials

A more immediate use is for lesson planning and materials development. Voyant Tools allows teachers selecting texts for use in class to analyze their level of difficulty as determined by lexical choices, sentence length, and vocabulary density. After choosing appropriate texts, teachers can implement a backward design process (Wiggins & McTighe, 2005) to prepare suitable assessments and rubrics focused on desired learning outcomes then

create engaging and effective lessons, tasks, and materials around the texts and tools in line with DDL.

Learning

Voyant Tools could promote learning following a DDL approach. While teacher-centered classes are possible, educators interested in student-centered, discovery or experiential learning paradigms may prefer to have learners use different devices to play with the interactive tools whether for autonomous or collaborative learning. The software affords users opportunities to explore texts for multiple purposes and from a variety of angles, so teachers could devise any number of activities, from simple language exercises to complex projects.

An example activity for intermediate or above learners could involve integrated reading-writing skills. In this case, teachers could have students use Voyant Tools to read, or interpret, texts of their choice or assigned materials by doing distant reading. This way, they could use the software to get the gist of the content. Next, teachers could design a scavenger hunt task for students to find the main theme or key details about people, settings, or events in readings. In addition to making these discoveries, learners could practice noticing skills by studying texts for discourse-related (genre, mode, and register/style) characteristics. While examining charts, graphs, and tables for instances of field-related vocabulary or features of spoken or written, formal or informal, polite or impolite, or gendered language, they could simultaneously develop their multimodal literacies. Meanwhile, using the software could promote critical thinking especially as choices have to be made about inputs and tools, and the visualizations are open to interpretation.

Following up with a writing task to practice productive skills, teachers could ask students to use the visualizations as prompts for text reconstruction or, more creatively, composition. After writing a draft, learners could use the software for self- or peer-review activities. Analyzing and comparing their own writing or that of their peers for patterns (e.g., in word frequency/repetition, (non)target vocabulary, L1 influence, or simple/complex structures) could lead to more systematic revision. Finally, they could enhance their technological ease and proficiency by taking advantage of the embedding capabilities of Voyant Tools and including exports in any web-based reporting. By using Voyant Tools, students can simultaneously learn about content, gain language awareness, hone language skills, and develop other essential 21st century skills.

Summary

Here, I have given a description and evaluation of Voyant Tools, its inputs, display outputs, and export options, and discussed some applications related to DDL. This eclectic collection of free tools, which is unfamiliar and underutilized in English language education, is a handy resource that can quickly and automatically analyze authentic and pedagogical texts for trends and patterns for a variety of purposes (language research, teacher training and professional development, and classroom teaching and learning) and audiences. A weakness is that the software may not be ideal for projects requiring huge corpora or part-of-speech tagging, and as for most activities involving technology, some training and careful task setup could contribute to more effective use. Hopefully, this does not dissuade anyone from experimenting and doing more work with the software in the future.

References

- Ananiadou, S., Thompson, P., Thomas, J., Mu, T., Oliver, S., Rickinson, M., Sasaki, Y., Weissenbacher, D., & McNaught, J. (2010). Supporting the education evidence portal via text mining. *Philosophical Transactions of the Royal Society A*, 368, 3829–3844.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. New York: Cambridge University Press.
- Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O’Keeffe, & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359-370). London: Routledge.
- Pérez-Paredes, P., Ordoñana Guillamón, C., Van de Vyver, J., Meurice, A., Aguado Jiménez, P., Conole, G., & Sánchez Hernández, P. (2019). Mobile data-driven language learning: Affordances and learners’ perception. *System*, 84, 145-159.
- Sinclair, S., & Rockwell, G. (2020). Voyant Tools, v. 2.4 (M29). Retrieved from <https://voyant-tools.org/>
- Walsh, S. (2011). *Exploring classroom discourse: Language in action*. New York: Routledge.
- Warschauer, M., Yim, S., Lee, H., & Zheng, B. (2019). Recent contributions of data mining to language learning research. *Annual Review of Applied Linguistics*, 39, 93-112.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd Ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Willis, D. (2003). *Rules, patterns and words: Grammar and lexis in English language teaching*. Cambridge: Cambridge University Press.

About the Reviewer

Jocelyn Wright, Associate Professor at Mokpo National University, has been teaching English in South Korea since 2007. Her interdisciplinary background is in education and linguistics, and she is interested in many areas, including critical curriculum, teaching methods, and materials development. She discovered Voyant Tools after taking a doctoral course on sociometry.

<jocelyn@mokpo.ac.kr>

© Copyright rests with authors. Please cite TESL-EJ appropriately.