

# Visualisation Approaches for Corpus Linguistics: towards Visual Integration of Data-Driven Learning

Jonathan C. Roberts\*  
School of Computer Science  
Bangor University, UK

Ana Frankenberg-Garcia†  
School of Literature and Languages  
University of Surrey, UK

Robert Lew‡  
Faculty of English  
Adam Mickiewicz University, Poland

Geraint Rees§  
School of Literature and Languages  
University of Surrey, UK

Nirwan Sharma¶  
School of Computer Science  
Bangor University, UK

## ABSTRACT

The compilation and use of corpora is not solely for research in linguistics. Among many other practical applications, corpora can be used to inform dictionaries, grammars and syllabuses. They can also help language users directly by providing concrete examples of common practices and good examples. Data-Driven Learning (DDL) describes situations where tools and techniques of corpus linguistics are used to learn a language or a particular type of language. However, DDL has remained largely confined to the research community. Consequently, there is a need to better integrate corpora with language pedagogy, develop visual techniques that will enable DDL to be used by wider audiences, and explore how visualisation could help make DDL more integrated and interactive. This paper addresses this question by exploring how visualisation approaches for corpus linguistics can enhance DDL, with particular focus on improving academic writing.

**Keywords:** Information visualisation, digital humanities, linguistics, academic writing.

**Index Terms:** H.5.2 [Interfaces and Presentation]: User Interfaces—Graphical User interfaces (GUI). K.3.2 [Computing Milieux]: Computers & Education—Computer Science Education

## 1 INTRODUCTION

Writing can be a cognitively-demanding task. Writers have to organise their thoughts, translate them into words, and present texts in a way that conforms to the expectations of their readers. For example, to write a recipe, you need to know not just the names of the ingredients, but also that readers expect you to list all the ingredients and quantities first, before explaining how to mix them in a particular way and order. You also need to know how words are used together in recipes (for example, you do not *shred onions*, but you can *chop* or *slice* or *dice* them). And you normally give instructions in the form of imperative sentences like *Heat the oven to 180C, Line a 22cm square tin with baking parchment*, and so on. Likewise, to write a research paper, you need to plan what you want to say, translate your ideas into the words conventionally employed in your discipline, and put those words together in sentences. You also have to follow the structure of academic articles, which normally have a title, the names of the authors, an abstract, an introduction, one or more middle sections, a conclusion and a list of references. The more research papers (or recipes) you read, the more familiar you become with how they are written, and learning ends up taking place incidentally.

\*e-mail: j.c.roberts@bangor.ac.uk

†e-mail: a.frankenberg-garcia@surrey.ac.uk

‡e-mail: rlew@amu.edu.pl

§e-mail: g.rees@surrey.ac.uk

¶e-mail: n.sharma@bangor.ac.uk

two functions . It **provides** a user **interface** and controls the processor 's hardware . Microsoft 's Windows **provide** an **interface** that makes use of a WIMP ( win the hardware and **provides** a user **interface** . Modern operating systems include decision procedures **provide** a consistent user **interface** across applications . Moreover , systems ranging from **providing** a user **interface** to managing the memory subsystems to decision procedures **provide** an **interface** between the communicative and a decision procedure **provides** an **interface** between the dialogue protocol and resources by **providing** a well-defined **interface** to the resources . Any set of resources **provide** a complete management **interface** for buildings . It enables the user to resolve internal disputes , **provide** security , **interface** with the Lebanese government

Figure 1: A listing of the noun *interface* occurring with the verb *provide* in the Oxford Corpus of Academic English via the SketchEngine.

Corpus linguistics can help to accelerate this learning process by providing learners with a concentrated exposure to examples of texts, in what has come to be known as Data-Driven Learning (DDL) [3, 8, 16]. This is exemplified in Figure 1, which illustrates how a student of computer science could scroll down a corpus of computer science texts to find out how the verb *provide* and the noun *interface* have been used together in such texts.

Although DDL was first put forward more than thirty years ago, even today most writers know little about corpora and about how to engage with them to produce better texts [7, 10]. Our vision is to bring the tools and resources of corpus linguistics to writing environments through the development of visual techniques that will enable DDL to reach wider audiences. The solutions need to engage with users, allowing them to improve their writing through examples and suggestions from corpora, without distracting them from translating their thoughts into words. We believe that visualisation research can substantially help to this goal. We encourage researchers to take up this challenge, and believe strongly that interdisciplinary teams are the way to solve these challenges.

This paper lays the foundation of this challenge. The work presented here was undertaken as part of ColloCaid, a collaborative project funded by the UK Arts and Humanities Research Council involving researchers skilled in computer science, information visualisation design and human computer interaction, applied linguistics, corpus linguistics and pedagogical lexicography (www.collocaid.uk). After some background and related work (Section 2), we provide a case study of DDL within the visualisation domain (Section 3), offer a brief description of ColloCaid (Section 4), and discuss research opportunities in this domain (Section 5).

## 2 BACKGROUND AND RELATED WORK

There are many interactive computer language tools that can help writers. For example, electronic dictionaries, machine translation engines, spelling and grammar checkers, and even tools that report on the readability of written texts are now in the hands of every computer user. These systems can help authors produce better texts.

However, while some tools require writers to stop writing (and hence interrupt their thought processes) to look things up, other tools are only capable of providing reactive feedback to writers, showing ways to correct texts retrospectively. One of the main problems of writing, though, is that writers are often unaware of the limitations of their own texts [6]. They cannot look up information they do not know they need, and correcting only what is wrong will not result in any improvement beyond that [12].

DDL has been researched since Johns’s [16] work on teaching learners to engage with corpus data directly. His vision places the user in the heart of the learning and discovery process. It is an inductive approach that encourages learners to notice and therefore understand how words are used in context. Corpus investigation tools allow users to analyse patterns of use in authentic texts. Users can apply quantitative techniques to decide which words are more appropriate, and they can retrieve concentrated samples of how words have been used in exemplary target texts. Teachers can use corpora to prepare exercises to help learners focus on specific language problems, and learners can engage with corpora directly whenever the need arises, gaining autonomy of their learning.

Although early work in DDL was restricted to the few institutions with the capacity to provide teachers and students with access to corpus data, nowadays there are many corpora that are easily available online, including the British National Corpus (BNC), the Corpus of Contemporary American English (COCA) and Sketch Engine for English Language Learning (SkELL) [9]. Moreover, advances by corpus linguistics researchers has meant that it is now much easier for end-users (whether they are teachers or learners themselves) to create personal corpora, which can be used to investigate real-world examples that apply to individual situations.

This has been a remarkable journey, from the early 1960s concordance programs run from punched cards, to concordance software of the 1990s such as AntConc [2] and WordSmith tools, to today’s server-based software tools such as SketchEngine [17, 18], Wmatrix [19] and CQPweb [14]. Yet despite the overall positive evaluation of DDL [3], current DDL tools suffer from several disadvantages and limitations, including:

1. corpus software is not particularly user friendly;
2. most learners are not aware of how corpora can be useful for them;
3. most learners do not know which corpus to use or how to build relevant queries;
4. learners can get distracted by the results they receive (and follow non-useful directions of enquiry);
5. learners can misinterpret the results, and come to the wrong conclusions;
6. corpus consultation takes time;
7. learners need to stop what they are doing to consult a corpus.

Our vision is hence to move forward the current state of the art, to make DDL more *integrated* and interactive, personalising the corpus experience, and to make the experience more *visual*. Current visualisation techniques applied to textual data use simple strategies, such as highlighting text, linking words, and using glyphs [21]. Many researchers have focused on *close* and *distant reading*, especially helping the reader to analyse written text (for a survey of close/distant reading visualisation techniques see Jänicke et al. [15]). Researchers have visualised corpus data, using tag clouds and frequency plots [1], discourse trees [26], dependency diagrams [4], and even parallel coordinate plots (PCP) [5]. While such strategies do provide an overview of the corpus data, they are not focused on learning, or helping users to improve their writing skills.

### 3 CASE STUDY: VISUALISATION OF CORPUS DATA

Let us consider an imaginary student (Jennifer) who is starting to write a visualisation paper. She may struggle to know how best to express certain phrases. On the other hand, her colleague Susan is a

Table 1: Showing relative frequencies of the word visualisation in the Vis6 visualisation corpus compared to the Open Access journals corpus (oAJ).

Word	Vis6 Corpora		oAJ Corpora	
	Frequency	Freq/mill	Frequency	Freq/mill
visualization	22,424	272.8	58,977	147.0
visualisation	340	41.3	5,962	1.8

more experienced author, who has been publishing in the community for several years, and implicitly knows the appropriate language. It is quicker for Susan to write the text, because she has learnt the domain-specific vocabulary and has developed skills to understand which words work together. Susan has developed her knowledge by reading articles and remembering phrases from other authors’ work. When writing the documents, both authors will have a set of questions in their mind about the best way to phrase the written work. The difference is that Susan will be able to quickly recall previous examples. For instance, as they write they may ask:

- “What other words can I use instead of *visualisation*?” (Q1).
- “Should compound words in visualisation be written as one word, with a space or a hyphen, e.g., ‘bar chart’, ‘bar-chart’ or ‘barchart’?” (Q2)
- “Do I use *explained by* or *illustrated in* the figure?” (Q3).

The answers to these questions are implicitly held in Susan’s mind, but we need to make them explicit such that Jennifer will understand them quickly too. For this case study, let us discover the answers to these questions from the visualisation literature. To answer these questions, we have created a bespoke corpus from visualisation texts (we name it Vis6). We include articles published in *Transactions on Visualization and Computer Graphics* (TVCG) between 2012 and 2017 that were presented at the IEEE VIS conference. This provides a 6-year view on visualisation articles, contains 632 documents, and provides a suitable corpus of visualisation texts for use to start to answer some of these questions. We build the corpus using SketchEngine [17] and the English Penn Treebank part-of-speech tagset. It contains 6,303,737 words (in total), and has a lexicon of 134,663 (different) words. To compare results from the visualisation literature to other texts we also build another corpus. Our second corpus is created from the Open Access Journals (doaj.org) that contains over 2.6 billion words from 2,971,481 articles. While there is a size difference, we are able to compare normalised frequency counts, and metrics such as the logDice score [17], which allows comparison between corpora, and enables us to explore examples from other domains. The oAJ corpus was chosen because it also contains research papers, but from a broader set of subjects (including science, social science, medicine and humanities).

While we run through this brief case study, do remember that the aim of this case study is not to discover the answers to these questions per se, rather we want to demonstrate and emphasise the process and the appearance of the results. We will need to build a suitable Graphical User Interface to display the results and allow users to explore the data. We explain our current implementation towards this goal in Section 4, and then discuss wider visualisation issues in 5.

#### Q1. Alternative words to “visualisation”.

We could imagine Jennifer wanting to write about a new visualisation design that she has created. To improve her English, she knows that she does not want to use the word “visualisation” repeatedly in a sentence, so she wants to understand synonyms that would be suitable in this domain.

To explore this question, we created a thesaurus of words on both corpora (Vis6 and oAJ) using SketchEngine, which we use to investigate synonyms “visualisation”. Table 1 shows the frequency of the words visualisation in the corpora, Table 2 shows the top sixteen words and the word cloud in Figure 2 visualises the top 60

Table 2: Frequency of occurrence, showing the top 16 thesaurus words of “visualization” in Vis6 (left) and oAJ (right).

Word in Vis6	LogDice	Freq.	Word in oAJ	LogDice	Freq.
analysis	0.424	16043	characterization	0.365	152184
datum	0.401	31974	quantification	0.348	133989
view	0.378	11440	modeling	0.332	197221
technique	0.362	9869	identification	0.326	378283
system	0.361	11014	imaging	0.309	284370
model	0.358	11859	verification	0.306	61796
method	0.341	10291	inspection	0.302	61897
representation	0.341	5283	segmentation	0.300	90461
design	0.337	10760	monitoring	0.295	224880
task	0.325	12434	interpretation	0.292	200978
approach	0.312	10056	validation	0.292	194384
interaction	0.306	8791	determination	0.290	243256
information	0.306	14111	tracking	0.286	61982
result	0.298	10905	mapping	0.284	184834
feature	0.298	8635	computation	0.284	138676
graph	0.295	9205	exploration	0.283	60212

Table 3: Five compound words that are commonly written in visualisation articles; preferred word forms are highlighted.

Lemma	Freq/mill	Lemma	Freq/mill	Lemma	Freq/mill
barchart	8.02	bar-chart	7.54	bar chart	120.39
linechart	0.36	line-chart	0.24	line chart	46.21
scatterplot	161.61	scatter-plot	0.97	scatter plot	11.5
timeline	87.43	time-line	0.36	time line	7.17
treemap	91.56	tree-map	4.98	tree map	3.53

words. From this analysis we clearly see that there is a substantial difference between the thesaurus words in the visualisation corpus, compared to the general one. There are many more occurrences of visualisation in the Vis6 corpus, and there is only one intersecting word in the top 60 words of both corpora.

Several visualisation techniques could be used to display these results to Jennifer. Either the tabular form (Table 2) or word cloud visualisation (Figure 2) may be suitable. The exploration technique of linked views [20] could be used to enable juxtaposed view information [13] (for example, a word cloud in one view to be linked through highlighting to text editor [25]). In fact, from the word cloud, we can see that Jennifer may use phrases such as “shown in the visualisation”, or “shown in a view” or “shown in the technique”, etc. However, both views contain a lot of information for Jennifer to grasp, and it is unclear which results are better for her to use. Furthermore our goal is to integrate the visualisations in situ with Jennifer’s text, and develop explicit visualisation designs [13].

## Q2. Do authors prefer compound words (in the visualisation corpus) in their closed, hyphenated or open form?

In our scenario, we can readily imagine Jennifer wanting to discuss different chart types. Many of these visualisation forms are compound words, such as barchart, linechart, scatterplot, timeline and treemap. There are three ways to write compound words, either they are written without a space, with a hyphen, or with an added space. Table 3 shows statistical results for five words. While each of the variants are used in the visualisation corpus some are more common; for instance “bar chart” is more common than “bar-chart”, while scatterplot (as one word) is more frequent than other variants. There are several ways to visualise this data. While highlighting text is one way to visualise this data, it may be better to show alternatives, and also visualise the uncertainty of the data rather than focus on one answer. What is required is an in-depth design study [22, 23] into in situ visualisation techniques.

## Q3. “explain” or “illustrate”?

It is easy to imagine Jennifer needing to describe the content of her

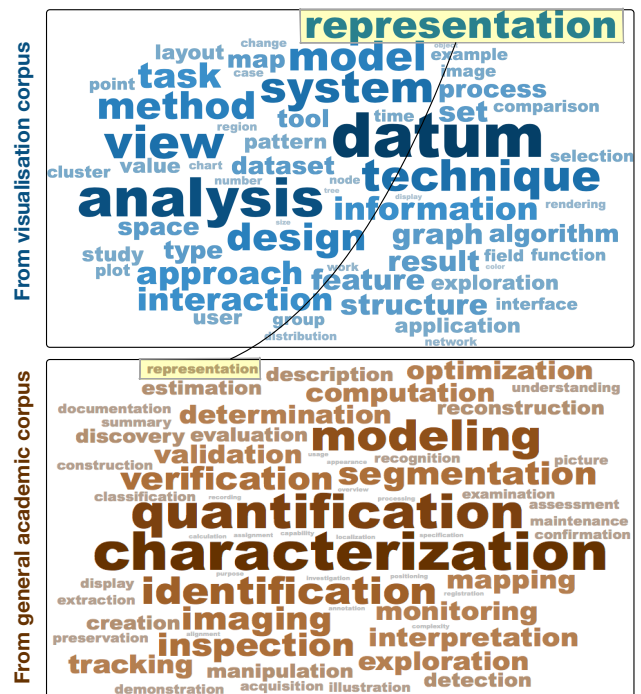


Figure 2: Word clouds of the top 60 synonyms of “visualisation”, showing stark differences between the visualisation corpus, Vis6 (top), and general academic corpus, oAJ (bottom). The only intersection is the word “representation”.

Table 4: Preferences for using *explain* versus *illustrate* by visualisation authors.

Favoured text	less favoured version
explain below	illustrate below
first explain	first illustrate
not explain	not illustrate
visually illustrate	visually explained
illustrate empirically	empirically explains
explained in Section x	illustrated in Section
illustrated in Figure x	explained in Figure

Figures. But what phrases does she use? Does she use *explain* or *illustrate*? By analysing our Vis6 corpus we can explore the preferences of phraseology of the visualisation authors. There are many domain specific variations and preferences of phrases that we have discovered. For succinctness, in this paper we highlight a few examples (see Table 4). For instance, we report that visualisation authors are more likely to say “briefly explain” than “briefly illustrate”, and have a preference for “illustrated in Figure x” rather than “explained in Figure x”. Similar to our previous questions, what is needed are better visual solutions to visualise this data to the user, in a way that encodes uncertainty and displays it in situ with the author’s text.

## 4 COLLOCAID: COLLOCATIONS FOR ACADEMIC WRITERS

Unlike the case-study presented in the previous section, which exemplifies corpus uses that are specific to the visualisation community, the ColloCaid project is aimed at helping novice users of academic English in general [12]. It harnesses data from the Oxford Corpus of Academic English (70 million words) and other corpora of expert academic writing to assist users as they write, in an integrated and interactive DDL approach. One of the distinguishing features of our tool is that the corpus data is invoked from a text editor, so

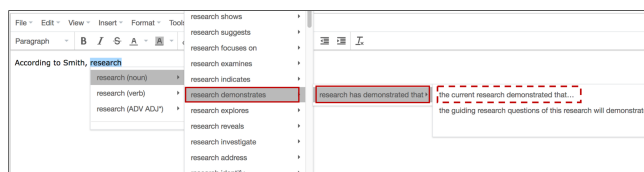


Figure 3: Screenshot of the ColloCaid demonstrator, showing how a user can select “research” to see different collocations and different examples.

that writers do not have to interrupt their writing to consult external resources. Opportunities for DDL are provided without users having to learn to use corpora, and the corpus data is curated to facilitate intake. Our focus is on the use of academic collocations, i.e., the ways words combine in the academic English [11]. For example, a student who is in the middle of writing a paper and wants to find a verb other than *do* that goes well with *research* can obtain this information intuitively and interactively. ColloCaid will suggest words like *carry out/conduct/undertake* and corpus examples of how these words combine.

The challenge is to develop ways of presenting this type information in a visually seamless way. We have started developing our prototype as a custom plugin for TinyMCE text editor. TinyMCE is a lightweight and extensible editor that is compatible with different browsers and operating systems and has widespread use on other online applications. Moreover, it provides rich text editing features with buttons that would be familiar to users of other word processing editors (such as Microsoft Word, Google Docs, OpenOffice etc), thereby making it easier for writers to use.

Figure 3 illustrates our plugin. Users can type sentences and paragraphs, and edit their text. We have integrated the database of collocations that we are developing [12]. Collocations prompts are highlighted automatically so that users know further information is available. They can choose to ignore and can carry on writing, or they can click on a keyword to explore possible alternative collocations that appear in drop-down menus. If users need further support to decide which collocation to use, they can then see corpus examples. For instance, the user could type “research” and flexibly obtain as much or as little information about the word. The last interaction shows lexicographically curated corpus examples. One example that appears is “research has demonstrated that”. The user can then select this text (through a double-click) to add it to their written words.

## 5 RESEARCH OPPORTUNITIES: VISUALISATION WITH LINGUISTICS AND DDL

From our case study (Section 3) it is clear to see how we can visualise corpus data to help users understand that some phrases are better. From the related work (Section 2) we also explained that there are many visualisation techniques that can be applied to corpus data; for example, in this paper we have included a word cloud visualisation of thesaurus data (Figure 2), highlighted text, alternative example texts. There are also many metrics that we have available: raw word frequencies, relative frequencies in words per million, association scores that quantify the degree of co-occurrence of two words (such as the logDice), etc. These can be used to organise potentially useful words for visualisation to benefit the writer. We use a weighted per-million frequency count to size words in the word cloud, Figure 2.

However, as highlighted in the background and related work (section 2) most of these visualisation techniques operate separately from the specific text that a user is writing. By contrast, our prototype (shown in Section 4 and in Figure 3) is an in situ solution. Currently, we use simple highlighting to select the nodes which the user can explore. We use drop-down menus, which allow the user to investi-

gate alternatives. We use association scores to order the alternatives, so that the better examples are listed first. This in situ editor has the potential for users to learn from these examples without interrupting their primary activity (a constructivist approach). In order to achieve our goal of developing a Visually Integrated Data-Driven Learning tool, we believe that there are three broad challenges to overcome.

**Challenge 1. How can we visualise the multivariate corpus data, such that it appears in situ and on demand, and that it encodes the necessary uncertainty of written language?** Especially, how can we develop solutions that appear as the author writes, and that they realise the context of the phraseology, and updates their work based on best practice (from corpus data)? The data contains metrics, words, phrases, structure and hierarchies, word types, etc., that can help writers if adequately visualised. However, it is not a matter of merely visualising the corpus data; rather, the visualisations need to be personalised to the user, and to phraseology that is suitable for their purpose. There are several research questions to address, including: What well-known visualisation techniques will be useful for the user? What new visualisation design ideas can be created to visualise this data? How can we visualise this multivariate data in situ? How can visualisation help with collocated words? How can visualisation help users realise that they have already made the same mistake  $n$ -times before?

**Challenge 2. How can we visualise text such that the user can explore alternative solutions?** Word exploration is very important. We want users to explore and interact with the system such that they serendipitously discover new words. There are still many research questions to ask, and much potential to making the edited corpora more visual. What is the best way to visualise top words of the domain? How can a user both explore different words and relate them to what they want? When is it suitable for the tool to be silent, or to proactively suggest words? When is it appropriate to launch a training tool, and take the user into a set of practical exercises? What level of intelligence is needed in the text analysis? How can the user control the system for their purpose?

**Challenge 3. What visual strategies will help users learn new words, and improve their writing skills?** Explanatory visualisation is a field that has a long history in education, but is less researched [24]. It is a subject that helps users to learn, and understand why something happens in a certain way. Research needs to take place into explanatory visualisation to show good practices in writing. What solutions can we invent that will visually explain words and phrases or collocations? How can the system visually explain problems (and solutions) in situ? Can visualisation be used to display the structure of the text, and help users improve their structure? How can we visualise text readability? How can we visualise the data such that users learn from their mistakes?

## 6 CONCLUSIONS

There has been much research into corpus linguistics, data-driven learning and corpus visualisation. While there is evidence of data-driven learning being used with some degree of success in the classroom, there are many opportunities to bring these ideas out of the classroom. We are now at a crossroads where all these technologies can be mixed together to create more visual and interactive data-driven learning systems, that will help users write more idiomatically, interact with their written work, learn good writing skills, and improve their written texts.

## 7 ACKNOWLEDGEMENTS

We acknowledge the Arts and Humanities Research Council, UK (AHRC) AH/P003508/1 for funding ColloCaid: combining learner needs, lexicographic data and text editors to help learners write more idiomatically.

## REFERENCES

- [1] A. Abbasi and H. Chen. Cybergate: A design framework and system for text analysis of computer-mediated communication. *MIS Quarterly*, 32(4):811–837, 2008.
- [2] L. Anthony. A critical look at software tools in corpus linguistics. *Linguistic Research*, 30(2):141–161, 2013.
- [3] A. Boulton and T. Cobb. Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2):348–393, 2017.
- [4] C. Culy, V. Lyding, and H. Dittmann. XLDD: Extended linguistic dependency diagrams. In *International Conference on Information Visualisation (IV)*, pp. 164–169. IEEE, 2011.
- [5] C. Culy, V. Lyding, and H. Dittmann. Structured parallel coordinates: a visualization for analyzing structured language data. In *Proceedings of the 3rd International Conference on Corpus Linguistics, CILC-11*, 2011.
- [6] A. Frankenberg-Garcia. Providing student writers with pre-text feedback. *ELT Journal*, 53:100–106, 1999.
- [7] A. Frankenberg-Garcia. Raising teachers’ awareness of corpora. *Language Teaching*, 45(4):475–489, 2012.
- [8] A. Frankenberg-Garcia. The use of corpus examples for language comprehension and production. *ReCALL*, 26(2):128–146, 2014.
- [9] A. Frankenberg-Garcia. *Corpora in the classroom*, pp. 383–398. Routledge, 2016.
- [10] A. Frankenberg-Garcia. Combining user needs, lexicographic data and digital writing environments. *Language Teaching*, 2018. doi: 10.1017/S0261444818000277
- [11] A. Frankenberg-Garcia. Investigating the collocations available to eap writers. *Journal of English for Academic Purposes*, 35:93 – 104, 2018. doi: 10.1016/j.jeap.2018.07.003
- [12] A. Frankenberg-Garcia, R. R. Lew, J. C. Roberts, G. Rees, and N. Sharma. Developing a writing assistant to help EAP writers with collocations in real time. *ReCALL*, 2018. doi: 10.1017/S0958344018000150
- [13] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011. doi: 10.1177/1473871611416549
- [14] A. Hardie. CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3):380409, 2012. doi: 10.1075/ijcl.17.3.04har
- [15] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. In *EuroVis – STARS*. The Eurographics Association, 2015.
- [16] T. Johns. Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal*, 4:1–16, 1991.
- [17] A. Kilgariff, V. Baisa, J. Bušta, M. Jakubíček, V. Kovvář, J. Michelfeit, P. Rychlý, and V. Suchomel. The sketch engine: ten years on. *Lexicography*, pp. 7–36, 2014. doi: 10.1007/s40607-014-0009-9
- [18] M. Kunilovskaya and M. Koviagina. Sketch engine: A toolbox for linguistic discovery. *Journal of Linguistics*, 68(3):503–507, 2017.
- [19] P. Rayson. From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13:519–549, 2008. doi: 10.1075/ijcl.13.4.06ray
- [20] J. C. Roberts. State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In G. Andrienko, J. C. Roberts, and C. Weaver, eds., *Proc. Int. Conference on Coordinated & Multiple Views in Exploratory Visualization (CMV2007)*, pp. 61–71. IEEE Computer Society Press, July 2007.
- [21] J. C. Roberts, A. Frankenberg-Garcia, R. Lew, and G. Rees. Visualisation and graphical techniques to help writers write more idiomatically. In *Poster, proceedings IEEE VIS 2017*. IEEE, New York, NY, USA, 2017.
- [22] J. C. Roberts, C. Headleand, and P. D. Ritsos. Sketching designs using the five design-sheet methodology. *IEEE Transactions on Visualisation and Computer Graphics*, 22(1):419–428, Jan 2016. doi: 10.1109/TVCG.2015.2467271
- [23] J. C. Roberts, C. J. Headleand, and P. D. Ritsos. *Five Design-Sheets – Creative design and sketching in Computing and Visualization*. Springer, 2017. doi: 10.1007/978-3-319-55627-7
- [24] J. C. Roberts, P. D. Ritsos, J. R. Jackson, and C. Headleand. The explanatory visualization framework: An active learning framework for teaching creative computing using explanatory visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):791–801, Jan 2018. doi: 10.1109/TVCG.2017.2745878
- [25] M. A. Wright and J. C. Roberts. Click and brush: A novel way of finding correlations and relationships in visualizations. In L. Lever and M. McDerby, eds., *Proceedings Theory and Practice of Computer Graphics*, pp. 182–196. Eurographics Association, June 2005.
- [26] J. Zhao, F. Chevalier, C. Collins, and R. Balakrishnan. Facilitating discourse analysis with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2639–2648, Dec 2012. doi: 10.1109/TVCG.2012.226