

Co-creating an online learning environment to support academic writing: Lessons learnt in an interdisciplinary setting

N. Sharma*
Bangor University

G. Rees†
University of Surrey

P. W. S. Butcher‡
Bangor University

R. Lew§
Adam Mickiewicz University

A. Frankenberg-Garcia¶
University of Surrey

J. C. Roberts|| (Member, IEEE)
Bangor University

ABSTRACT

In this short paper, we discuss the user-centred design process in the development of an online learning environment for learners of English for Academic Purposes (EAP). The ColloCaid project is a research collaboration between researchers from Applied Linguistics and Lexicography, Human-computer Interaction and Visualisation to develop a learning tool which provides users of academic English language an online environment to provide real-time suggestions to improve the vocabulary and fluency of their texts. Although still being developed, our online environment has received great interest from a range of users interested in improving their academic writing. The collaboration has revealed design insights which may be of interest to the researchers interested in the development of interactive learning environments.

1 INTRODUCTION

To solve today's challenges, researchers must work together across disciplines. By consolidating skills and knowledge, and collaborating together researchers can discover novel ideas. It is often harder to gain funding for research across disciplines because the agencies are focused around one discipline, publication can be challenging with journals finding it difficult to locate appropriately skilled reviewers, and with different working patterns, expectations and traditions between disciplines developing excellence in research can add further challenges [13]. For many years, academics, researchers and funding agencies have been discussing these issues, asking how to embrace and facilitate collaborative and interdisciplinary research. For instance, Rhoten and Parker [20] write about the risks and rewards of following an interdisciplinary approach, in 2015 Nature published a special issue on the need for interdisciplinary research, and how the world's biggest problems can only be tackled with teams. While interdisciplinary work can be challenging at times, it is often rewarding for the researchers involved, indeed Ledford [16] writes that it is a necessity, writing "we have to bring people with different kinds of skills and expertise together, no one has everything that's needed".

In this short paper we present our experiences and lessons learnt when collaborating on ColloCaid at the intersection of visualisation and the digital humanities. This project has brought together linguists, lexicographers, Human Computer Interaction (HCI) researchers and visualisation experts, with an aim to develop a novel interactive learning environment utilising visualisation. HCI research

is interdisciplinary in nature as it borrows language and methodology from other scientific disciplines thus providing a suitable setting to design and develop novel interactive technologies [17]. The project is funded by the UK Arts and Humanities Research Council (AHRC); in fact, it is to be commended that each of the UK's research councils have an interdisciplinary approach, and that the AHRC specifically identifies modern language research as an area of strategic priority, especially encouraging interdisciplinary approaches (<http://ahrc.ukri.org/innovation/>).

After background and related work (Section 2), we present how we chose an Agile design methodology, and explain how the tool is being jointly designed, developed and evaluated (Section 3). We reflect on our choices, and work practices, and draw together final remarks and lessons learnt (Section 4).

2 BACKGROUND & RELATED WORK

While there are many digital dictionaries and grammar support programs, our focus on word collocations is novel. Words that combine with each other in a natural way, are said to be collocated. In fact, Krishnamurthy [14] defines word collocation as "lexical items occurring ... with a greater frequency than the law of averages would lead you to expect". For instance, in visualisation authors would write *barchart* in preference to **barplot*, and would write *pie chart* and not **pie plot* [21, 22]. Collocations are particularly challenging for writers less familiar with the language in which they are writing or with the type of text they need to produce.

Collocations can be understood as co-occurrences of words, within a relatively short distance of each other in text, which are found more frequently than would be expected if word combination happened at random. Since the arrangement of words in language is non-random, collocation is a question of degree. For example, *black coffee*, a fairly conventional phrase, is a strong collocation, while the unconventional *dark coffee* is a considerably weaker one. Academic writers, particularly those new to the genre, frequently run the risk of breaking collocation conventions. The results range from the seemingly trivial, e.g. the use of more colloquial *do research* instead of the more formal option preferred in writing *conduct research*, to more serious errors e.g. **increase of pollution* instead of *increase in pollution*. In academic contexts, such breaks with convention impede the effective communication of knowledge.

Writing is a cognitively demanding process. Interactive systems can be developed to aid user learning by providing suggestions during the writing task. Writers need to be able to communicate their ideas effectively [7] and select words that convey the right meaning [24]. There are many tools that can help an author, from electronic dictionaries, grammar and spelling checkers, and even tools that analyse and report on the readability of the texts. Most of these tools rely on underlying data analysis, particularly corpus linguistics. There is a wide range of (mostly) server-side systems for corpus compilation and analysis including AntConc [2], SketchEngine [12, 15], Wmatrix [19] and CQPweb [9]. However, while these tools are powerful, time and resources are needed to teach users how to use them [5]. Furthermore most of these tools create corpora that are

*n.sharma@bangor.ac.uk

†g.rees@surrey.ac.uk

‡p.butcher@bangor.ac.uk

§rlew@amu.edu.pl

¶a.frankenberg-garcia@surrey.ac.uk

||j.c.roberts@bangor.ac.uk

separate to any editing and authoring that the user would perform. Subsequently, what is required is an integrated system, where users see the words they want to use in-situ with the texts that they are authoring, and be guided to best practices within the context of their writing [8].

In the case of ColloCaid, the language suggestions provided to writers are based on a database which is carefully curated by language researchers. It works by providing contextual suggestions based on what a user is currently writing in the text editor. Although similar information may be searched through online tools (such as dictionaries) users may not actively seek this information as they may not even realise that they can improve their existing text. Moreover, users may not want to interrupt the ‘flow’ of their writing by opening another window to search for a language doubt [10]. Learning in context is thus helpful, as a user may get suggestions when they are actually performing the task of writing. Moreover, providing triggers to use these suggestions is also imperative, as users need to be informed that suggestions exist for certain keywords. The main requirement of the tool is therefore to provide suggestions which are triggered when a user input keywords defined in a carefully curated linguistic database. Using this commonly defined aim in an interdisciplinary setting [1] we developed an online text editor. In this paper we provide a brief introduction to the dataset, the design rationale and decisions made, the resulting interactive tool and how it supports the academic writing tasks.

3 THE INTERDISCIPLINARY SETTING: SYNTHESISING VIEWPOINTS.

For ColloCaid, we took the decision early in the implementation phase to follow an **Agile development strategy** [4]. This is, perhaps, one of the most fundamental decisions of the project. We took the decision because (1) we wanted to have an early version of the tool to test, (2) we needed to work together as a team, while (3) also allowing change to occur in such a way that we can innovate and develop new ideas and novel research solutions. We felt that this strategy would work because we already had a strong vision from the proposal document, which gave us the initial requirements and development goals. This plan gave us a clear starting point. We agree with Highsmith and Cockburn [11] who write “Working through producing a plan drives the team members to think through their project and its contingencies”. We knew that we would need to develop a text editor, that would provide language suggestions to learners of academic English, that we would focus on word collocations and integrate word visualisation in the editor in-situ. We knew that we needed to create a bespoke lexicographical database and a text-editor to allow users to write sentences and paragraphs.

But we still had many research questions, including: How do we combine the editor with the lexicographic database? How do we visualise collocations? How do we visualise the word suggestions in-situ? There are many possible words to suggest; how do we display these possibilities? etc. Furthermore, with this interdisciplinary project we needed to bring together many skills. Best practices in lexicography and writing research, crafted and accurate development of the editor database, good practices in human interaction for the text editor, and visualising the words to the users. Furthermore, with the Agile method, we needed to test the software on users, and gain feedback such to improve the initial beta versions. How many users? What questions to ask the users? How to organise the evaluation sessions? etc.

Putting an appropriate suitable team of researchers together, is known as a challenge and important stage for any multidisciplinary research team. As a team of researchers, while we had not worked together, part of the team had known each other through other researchers and prior research. We got together because of our motivation to research in digital humanities and integrate visualisation techniques. This was a fortunate situation and definitely helped to

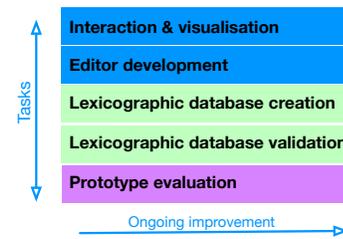


Figure 1: The main parts to the project, which have clear separation of goals and tasks.

create an interdisciplinary team with a wide range of skills. Furthermore, another aspect of the team was that as academic researchers we were interested in dissemination of our research through publications. We had similar expectations. In academic research, it is not only the outcome of the project and the development of the tool, but the presentation of the work, and demonstration of results. We are not saying that we do not have different opinions or work-practices differences to overcome, however we have a buy-in to the project and shared overarching research goals. Certainly, there are clear, and at times challenging, differences in work-practices, especially over publication strategies. For instance, there are many conferences and workshops (lexicography, corpus linguistics and computer-assisted language learning) throughout the year for language and linguistics researchers, whereas for visualisation there are fewer, yet larger conference venues. Most visualisation conferences review fully written papers, whereas the language and linguistics conferences review and select presentations on abstracts, before the authors write the journal articles, as is common practice with many disciplines. Another challenge was appreciating the expertise of each discipline and getting familiar with work-practices. For example, for the Human Computer Interaction (HCI) and visualisation researchers it was indeed a challenge to understand the importance of a carefully expert curated data-set of collocations as opposed to using a computer compiled corpus of millions of words and extracting suggestions algorithmically.

Additionally, there were sometimes challenges due to the location of researchers as the team is located across three sites (Surrey, UK; Bangor, UK; Pozna, Poland). However, we have held many (and regular) video conference meetings, along with several face-to-face workshops, to complete the next working prototype. This has meant we can entrust different sites to complete parts of the work independently, and report to the other collaborators through the video conference calls, and come together to complete the next prototype. Figure 1 shows a schematic diagram of the main parts of the project. This division also helped us to separate the workload, and take ownership of different tasks, at a given time.

The corpora. In our project we harnesses corpora, large collections of machine readable text, containing professionally published academic writing to curate data for the tool under development. The starting point, in order to provide collocation suggestions to writers, involves developing a set of collocation nodes, or a base word, which triggers the tool when written by the user. When compiling our list of nodes we consulted widely recognised studies of academic lexis to ensure that the bases were maximally useful to academic writers [8]. In our data ‘bases’ combine with ‘collocates’ to form collocations. The strength of these collocations is indicated by their ‘association score’, calculated according to Logdice statistics. This score determines the order in which collocations are displayed to the user in the tool, permitting users to see the strongest associations first. ColloCaid bases pertain to three parts of speech (POS): nouns, verbs, and adjectives. The bases form the keys to the lexicographic database underlying the tool. The notion of grammatical ‘relation’ is also important one. It describes the way in which a collocate

BASE	SET	PART	RELATION	COLLOCATE	COLLOC	LONGEST COMMONNES	CO-ASSOCIATION				
							NCE	SCORE	EXAMPLE1	EXAMPLE2	EXAMPLE3
approach	n	subject_of	proves			approach has proved	26	7.45	judging by the responses of	the benchmark approach	these sequential approache
approach	n	subject_of	allows			approach allows	53	8.07	deductive approaches allow	such an approach allows	such an approach may allow
approach	n	subject_of	involves			approach involves	54	7.78	the fourth approach involves	a more formal approach	a third approach involves sc
approach	n	subject_of	emphasizes			approach emphasizes	37	8.02	the first approach emphasizes	this approach emphasizes	one approach emphasizes h
approach	n	subject_of	yields			approach yields	21	7.45	such varying approaches will	this approach can yield	both approaches yield the s
approach	n	subject_of	provides			approach provides	96	7.53	nor did these approaches	valuable insights provided	such an approach provides ar
approach	n	subject_of	focuses			approach focuses on	37	7.74	both approaches focus on	our approach focuses on	the first approach focuses on
approach	n	subject_of	offers			approach offers	51	7.72	adopting the above approach	the advantages that both	the approach offers a novel n
approach	n	object_of	advocate			advocate an approach to	94	8.12	we have advocated this same	this statement advocates a	qualitative researchers advc
approach	n	object_of	follow			following the approach	210	7.9	we followed the traditional	it may be possible to follow	we follow this approach in e
approach	n	object_of	develop			develop an approach	313	8.24	transnational firms have	he also developed a unique	we must start to develop mi
approach	n	object_of	employ			employ an approach	137	8.1	they employed a systematic	another notable difference	we employed a maximum li
approach	n	object_of	adopt			adopt an approach	761	10.45	the various approaches	the country has adopted an	the theoretical approach adv
approach	n	object_of	take			take an approach	834	8.81	in these cases another	a different approach is	the funded projects took div
approach	n	object_of	apply			the approach applied	223	8.53	gene prioritization	the type of aspects	the only real experimental i
approach	n	object_of	use			the approach used	1,326	8.68	a feasibility study using a	this is an approach used	the disadvantages of using t
approach	n	modifier	holistic				171	7.45	this statement advocates a	an interesting strategy for	the tasks require a holistic a
approach	n	modifier	traditional				378	8.1	there are two traditional	a more traditional	we followed the traditional
approach	n	modifier	theoretical				304	8	the usefulness of a combined	debates between different	the second theoretical appn
approach	n	modifier	methodological				130	6.99	the standard methodological	a broad range of topics and	we also wish to present our
approach	n	modifier	integrated				168	7.32	management may take the	information derived from an	one quite common use of ar
approach	n	modifier	analytical				134	7.01	another common analytical	a framework analytical	there are many analytical ap
approach	n	modifier	alternative				488	8.63	an alternative approach using	an alternative approach	an alternative approach has
approach	n	modifier	systematic				151	7.1	a systematic approach to the	systematic approaches have	it requires us to develop a s
approach	n	preposition	to			(study/analysis/problem)	9,033	9.68	two distinct approaches to	the approach to genre	an engineering approach to

Figure 2: A sample database extract for the keyword *approach*

combines with a base. For example, in practice, a writer might have the noun *approach* in mind and wonder ‘What does an *approach* do?’ to answer this question the writer effectively, albeit several steps removed from this meta-language, accesses the collocates in the ‘subject of’ relation for *approach* and, perhaps, selects the collocate *emphasise* forming the collocation *approach+emphasises*. Similarly, a writer might wonder ‘Can I write *do an approach*?’ On this occasion the writer accesses the collocates in the ‘object of’ relation where *adopt + approach* and *use + approach* are revealed as good collocation options. A complicating factor is that bases, collocates, and association scores tell the user little about the arrangement, form of the words, or any intervening words contained the collocation. This is important since certain collocations exhibit preferences in these regards. For example, *use + approach* is typically realised in the passive voice as *approach used*. In our data, this information is encapsulated by the ‘longest commonest match’ measure; the longest most frequently occurring string containing any form of the base and collocate. Occasionally, writers might benefit from further information about the context in which a collocation is typically used. For this reason three extracts from the genuine academic texts in our corpora are provided as examples.

Text editor and visualisation. We identified an online text editor to be the most suitable platform to develop the prototype as the users can access it using their existing internet browsers thus eliminating the need to download additional software. An online editor would also be compatible across multiple devices and operating systems. We also identified that it would be useful to have the interface and features (formatting text, adding tables, inserting lists etc.) which closely match existing text editors in order to increase familiarity while providing users options to perform some basic functions during the writing process. In order to trigger suggestions for text it was agreed to use a mechanism similar to existing text editors (such as highlighting spelling mistakes), however the notifications may need to be less intrusive so that users may not perceive their existing text as mistakes which may need to be corrected using suggestions. To support in-situ text editing and user learning it was agreed that suggestions would be displayed as structured information in the form of interactive menu-items which when clicked would insert the suggestions in the text. The structure of the menus and submenus was identified from linguistic theory and praxis, hence providing a rational reasoning to assist user decision-making. Additional contextual information such as part-of-speech, syntactic relation etc. which users will need for decision-making may also need to be included in the menu and submenus. Considering the short-term memory limitations for information processing (7 ± 2 rule) it was also agreed to limit the suggestions to a maximum of eight per menu

to prevent overburdening the user [18].

Using these initial set of interface and information guidelines we prototyped the interface by developing a custom collocation suggestion plug-in for TinyMCE, an open source online rich text editor. TinyMCE was selected as it is widely used online with an interface which is similar to text editors such as Microsoft Word, Google docs etc. We customised the interface by extending the features of TinyMCE (see Figure 3A) and creating a ColloCaid plugin to our website. With the editor, the user can start typing any relevant text directly into the editor or copy paste formatted text from an existing document. The editor checks for collocation suggestions when the user presses spacebar, comma, backspace, return or tab keys on their keyboard while writing. The keywords, for which suggestions exist, are highlighted using a green dotted underline on the interface (as shown by the word “concerns” in Figure 3A). The users can choose to keep writing the text by ignoring the suggestions or click on highlighted keywords to see language suggestions. When the user clicks on one of the highlighted suggestions, an interactive menu (with sub-menu items) display the suggestions to users which users can insert into text by clicking on them. In the ColloCaid interface in Figure 3A, the user has first clicked on the *approach* keyword which was highlighted in text bringing up a menu item with suggestions. A mouse-over event on the first menu item, i.e., *approach proves*, opens a sub-menu of options for the user in the example.

We use round brackets to include additional contextual information for the menu items where it is appropriate such as identifying the keyword as a noun, verb or adjective, highlighting different senses of a keyword etc. This information is useful as it may help the user to easily make relevant decisions on which suggestions may be appropriate for their text for a given context. A ‘More’ menu button is displayed for keywords which have more than 8 suggestions to prevent scrolling and overburdening the user (for example this is shown under *approach offers* menu-item in Figure 3A). Clicking on the ‘More’ button displays additional suggestions on the right in a side menu. For the examples, it was decided to highlight the collocate and base keywords so that the user can easily comprehend the usage of the suggestions in the examples and apply them to their text. For example in Figure 3B three examples under the *approach involves* menu items are shown where both *approach* and *involves* are highlighted using bold text in the examples using bold text.

4 LESSONS LEARNT

In this paper we have described our design and interdisciplinary development process for the ColloCaid project. We have described how our collaboration has led to the development of the underlying

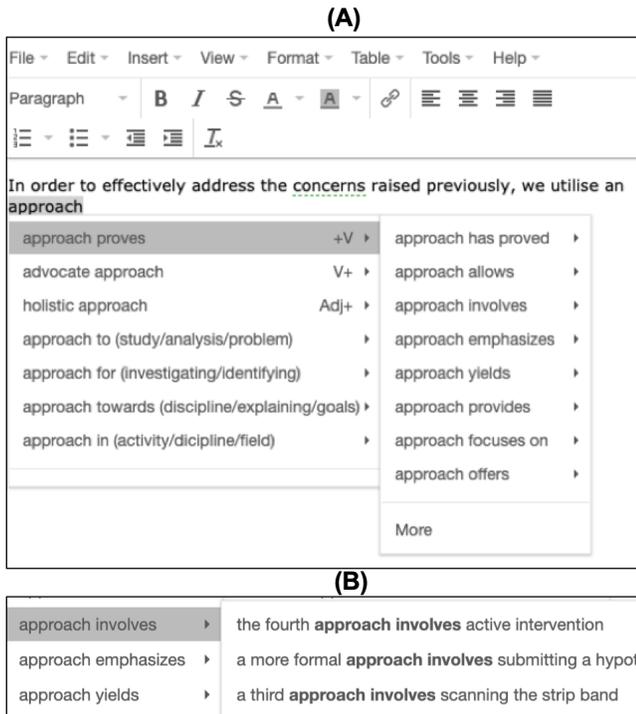


Figure 3: (A) The ColloCaid Interface showing the highlighted nodes or bases in the text which, when clicked by the user, gives suggestions in a layered approach through menu and sub-menu items. The inset (B) shows how further examples appear, to provide examples of the base in context with the relevant collocation highlighted in bold.

database and how we have developed the interactive editor, to help users to improve their writing. Reflecting on academic practice, not only helps us as researchers improve our practices, but can be shared to help other researchers perform better. The ColloCaid project is only one of the many interdisciplinary projects that team members have been involved with, but it provides a useful case study of interdisciplinary research. There are many potential lessons that we could discuss, such as using Agile methodology, having a strong vision, collaborating between two creative subjects, etc. But, for this paper we focus on three: (1) the degree of interdisciplinarity, (2) need for learning, and (3) shared visions and outcomes.

(1) Degree of interdisciplinarity. Aboelela et al. [1] discuss the degree of “cooperation or interaction” and of “communication and sharing”. How much do the disciplines collaborate, how often do they communicate, and how much they keep their own language (or create a common language)? There is a balance to be had. For our project, it was important to have a shared vision of the project. As a researchers we have a innate and common goal to publish our findings. In fact, our common vision document which was extremely useful to help us formulate overall goals for the project. Some of these goals helped the disciplinary teams to easily come to terms with certain issues such as the need for an interactive system which provides real-time suggestions to users. Similarly, we agreed to use different methods of evaluations for the system and mutually agreed to specific methodologies to evaluate the system.

We have regular communication using video conferencing, emails, face to face meetings to increase contact time between researchers. We additionally use technology mediated collaborative tools (such as OneDrive, Basecamp and Overleaf) as we have found them extremely useful for supporting our research and sharing information. All these activities and working styles we believe have led to greater

interaction and deliberation between researchers. Some of the issues such as different locations of working of the team we have tried to address by having more frequent meetings between postdoctoral researchers to address pressing issues and challenges or work together over days to finalise prototypes.

(2) Polymathy and the need for ongoing learning. Achieving creative polymathy-led research is difficult [23]. Development of a shared language and understanding can be argued as one of the outcomes of interdisciplinary research and to some extent, each of the researchers needs to learn about each others subject [1, 6]. Not to be experts in their fields, but to first develop enough knowledge such to have empathy for the decisions of researchers in other disciplines. Second, to be able to give presentations about the project with clarity, and third to give enough knowledge such to develop novel insight. We suggest that this learning should be ongoing, and consistent throughout the project. For example, the HCI researchers suggested that the System Usability Scale [3] could be used to evaluate usability, this was new to some of the team. While methodologies such as think-aloud and screen recording were mutually known. Learning, in ColloCaid, started at the start of the project, where we prepared and gave introductory presentations of each discipline. Furthermore, the HCI and visualisation team attended a language/linguistics workshop, and held our own project meeting directly after this event. This collaboration helped to coalesce the team and discipline knowledge within the team. But learning and collaboration is ongoing.

(3) Shared outcomes. One of the shared outcomes for researchers is the dissemination of findings. Although this project is still work-in-progress we have achieved some preliminary publications from the research collaboration. We have agreed to collaborate during writing publications and all researchers review and edit the papers, even when the paper that is being prepared is to be sent to a single discipline research journal. This we think helps us develop a shared language of research for the project and help in triangulation across disciplines [17]. One of the best forms of collaboration has been co-authoring papers such as this one, where the writing process helps us critically reflect on how to communicate discipline specific knowledge to an interdisciplinary audience. For example, in the current paper the background information on collocations and curated data-set sections have been written by language researchers in a manner suitable for a wider audience who may be less familiar with these concepts. In terms of outcomes we have also applied knowledge acquired from other disciplines in a novel manner back into our respective domains. For example, we have started to apply methodologies from one domain to the other, such as applying the System Usability Scale [3] from the HCI domain in the linguistic domain, and language corpus analysis techniques to the visualisation domain, resulting in a publication on word and collocation analysis in multiple views [21]. For future publications we have also identified journals and conferences which are more interdisciplinary in nature in order to communicate our findings to a wider audience than our respective disciplines.

In conclusion, this interdisciplinary project we found that formulating a shared understanding, through common goals, in the beginning and maintaining it during the research process was useful for developing mutual trust and cooperation between the researchers. The need for ongoing learning, whether through co-authoring publications or co-creating an interactive system can help create a collaborative setting where interdisciplinary research can further develop. This process of collaboration helps in building trust leading to shared ownership of the outcomes, tools and a common language of research, possibly leading to a transition from interdisciplinary to transdisciplinary research.

ACKNOWLEDGMENTS

This work is funded by AHRC grant AH/P003508/1, collocaid.uk.

REFERENCES

- [1] S. W. Aboelela, E. Larson, S. Bakken, O. Carrasquillo, A. Formicola, S. A. Glied, J. Haas, and K. M. Gebbie. Defining interdisciplinary research: Conclusions from a critical review of the literature. *Health Services Research*, 42(1p1):329–346, 2007. doi: 10.1111/j.1475-6773.2006.00621.x
- [2] L. Anthony. A critical look at software tools in corpus linguistics. *Linguistic Research*, 30(2):141–161, 2013.
- [3] A. Bangor, P. T. Kortum, and J. T. Miller. An empirical evaluation of the system usability scale. *International Journal of Human Computer Interaction*, 24(6):574–594, 2008. doi: 10.1080/10447310802205776
- [4] K. Beck, J. Grenning, R. C. M. amd Mike Beedle, J. Highsmith, S. Mellor, A. van Bennekum, A. Hunt, K. Schwaber, A. Cockburn, R. Jeffries, J. Sutherland, W. Cunningham, J. Kern, D. Thomas, M. Fowler, and B. Marick. Manifesto for agile software development. Agile Alliance, agilemanifesto.org, 2001.
- [5] A. Boulton. *Data-Driven Learning and Language Pedagogy*, pp. 181–192. Springer International Publishing, Cham, 2017. doi: 10.1007/978-3-319-02237-6_15
- [6] L. J. Bracken and E. A. Oughton. what do you mean?the importance of language in developing interdisciplinary research. *Transactions of the Institute of British Geographers*, 31(3):371–382, 2006.
- [7] 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
- [8] A. Frankenberg-Garcia, R. Lew, J. C. Roberts, G. P. Rees, and N. Sharma. Developing a writing assistant to help eap writers with collocations in real time. *ReCALL*, 31(1):23–39, 2019.
- [9] 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
- [10] A. Hasselgren. Lexical teddy bears and advanced learners: a study into the ways norwegian students cope with english vocabulary. *International Journal of Applied Linguistics*, 4(2):237–258, 1994. doi: 10.1111/j.1473-4192.1994.tb00065.x
- [11] J. Highsmith and A. Cockburn. Agile software development: The business of innovation. *Computer*, 34(9):120–127, 2001.
- [12] A. Kilgarriff, V. Baisa, J. Bušta, M. Jakubčėk, V. Kovvār, 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
- [13] J. T. Klein. Evaluation of interdisciplinary and transdisciplinary research: A literature review. *American Journal of Preventive Medicine*, 35(2, Supplement):S116 – S123, 2008. The Science of Team Science. doi: 10.1016/j.amepre.2008.05.010
- [14] R. Krishnamurthy. The process of compilation. *Looking up: An account of the COBUILD project in lexical computing*, pp. 62–85, 1987.
- [15] M. Kunilovskaya and M. Koviagina. Sketch engine: A toolbox for linguistic discovery. *Journal of Linguistics*, 68(3):503–507, 2017.
- [16] H. Ledford. Team science. *Nature*, 525(2):81, 1956.
- [17] W. E. Mackay and A.-L. Fayard. HCI, natural science and design: a framework for triangulation across disciplines. In *Symposium on Designing Interactive Systems: Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques*, vol. 18, pp. 223–234, 1997.
- [18] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.
- [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] D. Rhoten and A. Parker. Risks and rewards of an interdisciplinary research path. *Science*, 306(5704):2046–2046, 2004. doi: 10.1126/science.1103628
- [21] J. C. Roberts, H. M. A. Al-Maneea, P. W. S. Butcher, R. Lew, G. Rees, N. Sharma, and A. Frankenberg-Garcia. Multiple views: different meanings and collocated words. *Computer Graphics Forum*, 3 2019. doi: 10.1111/cgf.13673
- [22] J. C. Roberts, A. Frankenberg-Garcia, R. Lew, G. Rees, and N. Sharma. Visualisation approaches for corpus linguistics: Towards visual integration of data-driven learning. In *3rd Workshop on Visualization for the Digital Humanities*, at *IEEE VIS, Berlin, Germany*, October 2018.
- [23] R. Root-Bernstein and M. Root-Bernstein. People, passions, problems: The role of creative exemplars in teaching for creativity. In *Creative contradictions in education*, pp. 143–164. Springer, 2017.
- [24] V. Zakharov. Evaluation and combining association measures for collocation extraction. In *Proceedings of the International Conference IMS-2017*, IMS2017, pp. 125–134. ACM, New York, NY, USA, 2017. doi: 10.1145/3143699.3143717