A Blended System for Data-Driven Learning of English for Specific Purposes

https://doi.org/10.3991/ijet.v17i12.29653

Hengbin Yan
Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China
yhb@gdufs.edu.cn

Abstract—English for Specific Purposes (ESP) and Data-Driven Learning (DDL) are two constructivist and student-centered approaches to language pedagogy that are well-established in second language acquisition. Despite recent fruitful attempts to unify the two approaches for enhanced pedagogical effectiveness, current technological bottlenecks, characterized by a lack of specialized functionality and accessible interfaces for ESP learning, have limited their application in a broader range of scenarios such as blended language learning. To address these issues, the present paper proposes the design and development of a new DDL system tailored to the needs of ESP pedagogy, with functions specifically designed to foster student-centered learning and develop genre awareness at different stages of the ESP learning process. Built on a flexible modular architecture, the system utilizes state-of-the-art techniques in Natural Language Processing (NLP) for efficient multilayered linguistic annotation and indexing, made query-able through a user-friendly web interface that minimizes expertise required for DDL activities. We demonstrate the usefulness of the system with a case study showcasing its key functions for data-driven ESP learning, and discuss its potential use in an integrative blended learning environment.

Keywords—blended learning, English for specific purposes, data-driven learning

1 Introduction

1.1 Constructivist approaches to language pedagogy

Language pedagogy in recent decades has seen a gradual paradigm shift from teacher-centered to student-centered approaches whereby the student increasingly takes the role of an active *constructor*, instead of a passive recipient, of linguistic knowledge [1]. One important theoretical motivation behind the trend is constructivism, a theory of learning which has since become the dominant theoretical framework in language education [1], [2]. Constructivism postulates that learning only takes place when the student constructs mental representations of knowledge on their own by actively exploring, experiencing, interpreting and interacting with authentic input,

aided by proper use of tools, materials and activities [2]. Aligned with this tenet, constructivist approaches to language pedagogy have placed the student's individual cognitive development at the center of the language learning process, with the teacher playing the role of a facilitator and guide [3].

The constructivist theory has inspired and influenced the development of two important pedagogical approaches at the forefront of language teaching and learning: (1) English for Specific Purposes, which caters to students' specific, individualized needs and interests, providing approriate disciplinary contexts, materials and activities [4]; and (2) Data-Driven Learning, which enables students to construct contextualized knowledge through authentic input and inquiry-oriented activities, promoting cognitive development and learning autonomy [5], [6]. Despite their shared constructivist underpinnings, however, the two approaches have seldom been studied in conjunction within an integrated framework to reap potentially synergetic benefits.

1.2 English for specific purposes

English for Specific Purposes (ESP) is an approach to English language pedagogy explicitly directed for the specialized needs of language learners [7], [8]. One core assumption of ESP is that the problems in language teaching are unique to specific learners with specific needs and interests, characterized not only by disciplinary/genre variations in the subjects taught, but also differences in learners' motivation, language proficiency and learning styles etc. [9]. As such specificity can have a direct impact on learning outcomes, ESP practitioners have suggested that language instruction should be student-centered, avoiding the repetitiveness/redundancy in one-size-fits-all instruction [10]. ESP pedagogy should thus focus on the aspects of language that meet the needs of learners in specific contexts and be supported with materials that authentically reflect the target discourse community [9]. One important technique to accomplish this in ESP is *genre analysis* [11], where learners study the rhetorical structuring and linguistic patterning linked to disciplinary contexts of target genres (with *genre* commonly conceptualized as communicative events sharing communicative purposes recognized by the target discourse community [11]).

Despite the promises and numerous success stories of ESP [12], however, implementing the ESP paradigm in actual language courses remains a daunting task. Insufficient classroom sessions have limited students' exposure to ESP practice and exercises, making it difficult to profile individual needs and tailor instruction accordingly [13]. ESP teachers often lack specialized education and training on the specialist discourses of multiple disciplines and on effective identification of learner needs, which are central to effective preparation of ESP curricula [13], [14]. Moreover, as mass-produced textbooks seldom provide relevant, authentic and up-to-date materials for a target specialization, ESP teachers typically need to develop in-house course materials, collect authentic, specialized texts, and design relevant teaching tasks, all of which require extensive specialist and technical expertise and are challenging even for experienced teachers [9], [15], [16].

1.3 Data-driven ESP learning

To meet the demands of high-quality content-based instruction, ESP researchers have proposed the incorporation of Data-Driven Learning (DDL) [17], [18], where corpora are used to grant students access to authentic, dynamic resources in specialized disciplines [19], [20].

Aligned with the philosophy of ESP, DDL adopts a constructivist and student-centered approach to language learning [21], while mitigating some of its inherent problems. DDL aims to supply language learning with rich, authentic input through contextualized usage samples and frequency statistics on accessible interfaces, helping teachers and students to extract and prioritize common linguistic features important for specialized disciplines [20]. Teachers can promote students' discipline-specific genre competence by engaging them in corpus-based genre-analysis activities tailored to the target discipline, and provide targeted feedback on individualized problems [22], [23]. The burden of laborious teaching material preparation and compilation can also be eased, with the task shifting from the teacher having to provide all ready-made exercises and answers, to letting students immerse in data exploration guided by teachers. Through active corpus consultation and "exchanges" with the corpus texts, students learn to investigate and internalize not only specialized terminology but also essential types of genre-specific linguistic constructions beyond the word-level (e.g. phraseological patterns and collocations) known to be important to language learning [24].

However, when the DDL and ESP paradigms are deployed to actual language teaching curriculums, their implementation is constrained by several technological bottlenecks. Current DDL tools in use (e.g. Antconc [25], WordSmith [26]) are built for general-purpose corpus consultation and are not explicitly designed for the purposes of DDL or ESP. They thus do not incorporate functionalities that facilitate data-driven, ESP learning (e.g., functions to raise genre awareness via contrastive cross-genre statistics and visualization). Moreover, the core features available in these tools, such as concordance, are limited in function and difficult to learn and use. For instance, concordance lines are often chopped off, stripping users of necessary context [20], [21]. The query syntax can be hard for ESP teachers and students alike to master intuitively even after extensive training [21], [27]. In addition, while existing tools can retrieve concrete expressions (usually individual or contiguous sequences of words), they are often incapable of identifying more abstract and general linguistic constructions, such as *the* X-*er*, *the* Y-*er*, or Argument Structure Constructions [28], which entail the combination of multiple layers (lexical, syntactic, discourse etc.) of linguistic features and are difficult to capture using purely word-level sequences [29].

The problems above call for the creation of new software with improved, specialized functionalities. In the following sections, we present our work on building such software: a prototypical blended learning system [30] aimed at facilitating data-driven ESP learning. We describe the core features, architecture and frontend interface of the system, and discuss its potential blending into ESP classrooms.

2 Method

2.1 Core features and innovations

The proposed system is specifically designed for data-driven ESP learning activities, with specialized functions and interfaces to aid users in exploratory corpus-based learning (e.g. investigating cross-genre features and variation). It features two separate but related modes of exploration intended for users with different roles and expertise. The first is *free exploration mode* designed mainly for teachers and advanced users versed in the corpus query syntax. This mode allows these users to formulate, refine and store (groups of) reusable definitions of complex constructional patterns with accompanying descriptions. The second mode is the *guided exploration mode*, intended for students and those without sufficient technical knowhow. This mode does not require users to learn a corpus query language, which can be difficult and time-consuming even for experienced users [27], but instead allows them to select from a list of patterns predefined in the free exploration mode, applying the patterns to corpora of interest and optionally editing keyword slots in the patterns for specialized needs.

To ensure that the system can process large-scale corpora that represent authentic, dynamic language use, we use *Odinson* [31], a powerful and efficient information extraction framework for indexing and querying multiple layers of linguistic information (e.g. lexical, syntax, discourse). We utilize Odinson's flexible query language for precise definition and retrieval of complex patterns at varying levels of granularity, including more schematic/abstract patterns such as argument structure constructions.

The system features an interactive, web-based interface that allows for user customization, exposing parameters of backend features such as annotation, statistics and visualization, for maximized configurability (e.g. easy partitioning of corpora into contrastive segments). The system adopts data-driven mechanisms throughout its processing pipeline, from construction extraction, automated annotation, to statistical analysis. Extracted constructions can be used to differentiate genres by comparing their statistical distributions across genres or against a general reference corpus. A list of the most salient constructions are generated and visualized, ranked based on configurable criteria such as frequency, association measures, or distinctiveness across genres.

2.2 System architecture design

The system adopts a modular, pipelined architecture that groups core functionalities into independent modules communicating via defined interfaces (Figure 1.). This is to ensure the functionalities are easy to test, maintain and evolve, without being tightly coupled to low-level implementation details. The modules are designed to fulfill a number of functional requirements. First, pattern querying should be efficient (ideally in real time) so that users can query potentially large-scale data smoothly. Second, the query engine should allow the retrieval of flexible combinations of differ-

ent layers of linguistic annotation, so that a wider variety of important linguistic structures can be identified for study. Third, annotation of different layers of linguistic information on textual corpus data should be performed automatically to enable in-depth analysis. Fourth, statistics in visual formats should be provided for users to easily interpret the overall patterning underlying the data. Finally, the user interface should be designed to be accessible and easy to use, without requiring special technical expertise on the part of its users. The resulting system implemented in response to these requirements comprises the following six interrelated modules:

1. The corpus indexing module, built on top of Odinson, indexes corpus tokens and associated annotations, and enables efficient and flexible retrieval of linguistic patterns from multiple layers of annotations.
2. The database module provides a central storage for information about users, pattern definitions, corpus-related information (e.g. metadata that can be used to partition a corpus for comparative analysis), and the relationships among various types of data.
3. The annotation module is responsible for the automatic annotation of uploaded corpora. The default annotator provided is spaCy[1], an NLP framework with state-of-the-art syntactic analysis pipelines, but third-party processors can also be plugged in to attach layered (i.e. lexical, syntactic, discourse) annotations to the same corpus.
4. The management module is where the business logic resides and is responsible for managing and coordinating communication between the backend server and the frontend interface (e.g. decoding frontend user queries for backend pattern retrieval and processing).
5. The statistics module collects and aggregates results obtained from the stored corpora in response to user queries and computes detailed statistics to be presented on the web-based user interface.
6. The user interface module interacts with users and gives them easy access to configurable functionalities on a web browser, allowing them to execute various types of queries as well as perform statistical and visual analyses.

[1] https://spacy.io/

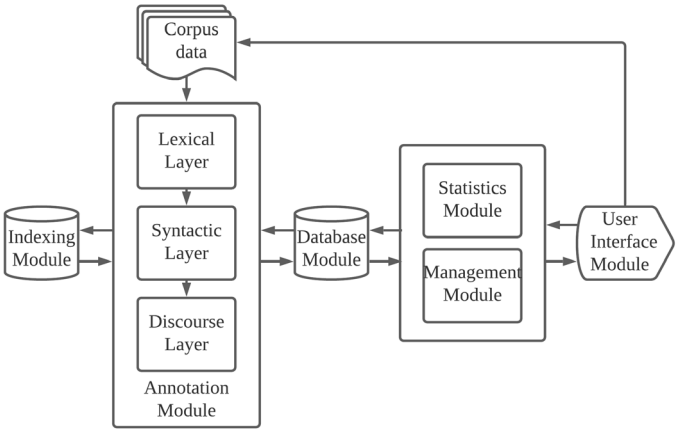

Fig. 1. Overall architecture of the system

3 Results and discussion

3.1 Web-based interface

The system interface developed in our prototype is composed of coordinated web pages designed to give users direct, flexible control in viewing and manipulating data (Figure 2.). The pages are used to organize groups of related functions and variables in a wizard-like fashion to guide users step by step, while affording them the freedom to switch between pages at any point in the analysis process.

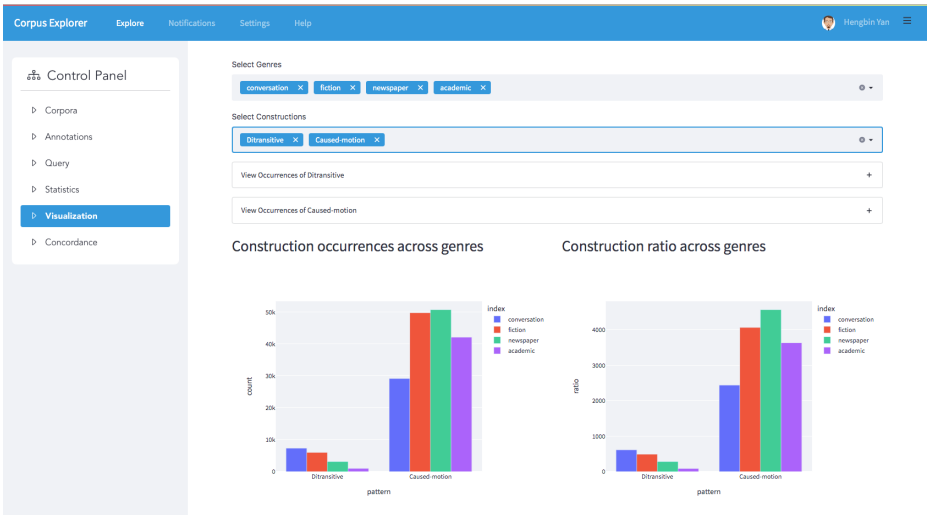

Fig. 2. A screenshot of the web-based user interface

The pages on the frontend are tightly integrated with the backend, which exposes variables through the Streamlit[2] framework to be configured by the user for specialized needs. The values in the frontend components (e.g. dropdown menus) are dynamically generated and specific to the selected corpus data. Any backend-related variables changed by the user on the frontend will be automatically submitted back to the server, with the affected components refreshed in response. Frontend options can also be configured dynamically with real-time updates on tabular and visual presentation, affording users the power to investigate results from multiple perspectives. Since most expensive computations are cacheable, the updates are performed efficiently.

3.2 Typical procedures from a user's perspective

Using the system for ESP learning, a user typically goes through the following steps for genre-based ESP analysis:

Corpus selection. The user first uploads a corpus to the system, or selects from a list of existing corpora. In the case of corpus uploads, a wizard guides the user in filling in the corpus metadata, e.g., the genre of the corpus, the time of compilation, etc. For contrastive analysis, more than one corpus can be uploaded/selected.

Corpus annotation. The user chooses the annotation layers (e.g. part-of-speech, syntactic relations) to apply to the selected corpora. The corpora will then be automatically annotated and indexed by built-in (powered by spaCy and Odinson) or third-party processors.

Corpus query. The user embarks on two possible modes of exploration into the selected corpora by defining target linguistic constructions with a set of rules using the query language, or by selecting from constructions previously defined by other users (e.g. course designers or instructors).

Result exploration. The user explores the results of the construction retrieval with relevant statistics, examining the constructional patterns at various level of granularity. The most abstract granularity shows statistics of all patterns as a group across different corpora. The next level allows the user to rank and filter patterns based on the statistics of individual patterns, including pattern frequency and individual pattern distribution across genres. Each pattern with its actual usage contexts can be viewed in a visual interface. Each matched component in the pattern can be individually highlighted and configured for statistical analysis.

Contrastive analysis. The user optionally selects different constructions for side-by-side cross-genre comparison. A number of statistical tests such as t-test, chi-square test, along with several visualization options, are available for analysis, allowing the user to directly discern the distinctive usage patterns and trends of different constructions across genres.

[2] https://streamlit.io/

3.3 Case study

In this section, we present a case study that demonstrates the process of data-driven ESP learning through cross-genre investigation of different types of Argument Structure Constructions (ASCs) (Table 1.), using the verb *give* as an example. ASCs are among the most well-known linguistic constructions on which numerous experimental and corpus-based studies have been conducted [32]. As an indicator of language users' language proficiency, they have important research and pedagogical implications [33]. The verb *give* is the prototypical and most frequent verb of the ditransitive construction, and is often the focus of investigation in language acquisition studies involving the construction (often in comparison with the to-dative construction, a subtype of the caused-motion construction) [34], [35].

While ASCs are found in various genres of texts and important for cross-disciplinary ESP learning, the cross-genre distribution of their different subtypes with specific verbs is seldom investigated. Without knowledge of such distribution patterns, learners formulating sentences with particular verbs (e.g. *give*) may find it difficult to decide on the use of different types of ASCs, e.g., ditransitive constructions (*somebody gives somebody something*) or caused-motion constructions (*somebody gives something to somebody*) appropriately in different genres of varying formality.

In this case study, we focus on demonstrating two core features of the system: (1) flexible query and retrieval of different types of linguistic constructions and (2) genre variation analysis through cross-genre comparisons of the statistical distributions of constructions. The corpus in use is a 4-million-word, genre-balanced sub-corpus (the Baby version) of the British National Corpus (BNC) [36], which represents the variety of English typically taught around the world to learners of English as a second or foreign language.

Table 1. Examples of Argument Structure Constructions (Adapted from [37])

ASC Type	Form	Meaning	Example
Transitive	Subj V Obj	X acts on Y	Tom kicked the ball.
Caused-motion	Subj V Obj Oblpath/location	X causes Y to move Zpath/location	Tom kicked the ball into the goalpost.
Ditransitive	Subj V Obj Obj2	X causes Y to receive Z	Tom kicked Jim a ball.
Resultative	Subj V Obj RP	X causes Y to become Zstate	Tom kicked the door open

Query and retrieval of ASCs. The subtypes of ASCs are defined using a set of rule-based criteria. For example, a ditransitive construction is defined as a construction with the following pattern: *Subj V Obj Obj2*, while a caused motion construction has the pattern of *Subj V Obj Oblpath/Location* [37]. While such definitions appear straightforward, retrieving the construction components (*Subj*, *Obj*, and *Obj2* etc.) from a corpus of non-trivial size requires the identification of syntactic relationships beyond the lexical level. For example, in the following sentence taken from the BNC, *Warm-hearted lions will give handicapped and underprivileged children from across the region a special day*, *Obj* (*children*) and *Obj2* (*day*) are separated by intervening words from *Verb* (*give*). In natural language, the distances between such components

can be arbitrarily long. As a result, these constructions cannot be reliably retrieved using existing tools that define query patterns linearly, such as AntConc and the COCA interface[3], without risks of producing erroneous outputs that skew the statistics and mislead users.

On the other hand, the syntactic annotations as well as the indexing mechanisms of our system allow the relationships among components to be concisely defined using a Semgrex-inspired syntax[4]. For instance, for the retrieval of ditransitive constructions, we can specify this rule: *{upos:VERB} ?>nsubj {} >dative {} >dobj {}*. Here, we distinguish *Obj* and *Obj2* using two labels from the annotation scheme of dependency grammar: *dative* (indirect object) and *dobj* (direct object). The system then retrieves relevant words with the specified relationships regardless of the linear distance between them. The query pattern, once defined, can be named, stored and reused in subsequent sessions (e.g., in the guided exploration mode).

Cross-genre investigation. Once the constructional patterns have been retrieved, the system computes cross-genre distributional statistics of each construction. Such distribution statistics are generally unavailable in existing DDL tools, with the exception of the COCA interface, which, although providing genre distribution statistics, has limited the statistics to single words and is incapable of showing pattern occurrences within a genre.

We compare the statistical distributions of the constructions in a visualization pane on the system. The results in Figure 3 reveal that, overall, the verb *give* is used more frequently (per 100,000 tokens) with the ditransitive construction than the caused-motion construction, appearing in the former more than three times as often as the latter. These results are unsurprising as they are corroborated by well-known evidence from previous corpus-based investigations on the constructional distribution of *give* conducted with smaller-scale data, which showed that *give* is the most frequent and distinctive verb of the ditransitive construction (whose central meaning of *transfer* matches that of the verb), with a much higher frequency and association strength than its caused-motion counterpart [38]. Beyond the frequency distribution of the two constructions, however, our system yields additional comparative statistics which demonstrate that the above distributional discrepancy holds across the four different genres, with the important exception of the academic genre. It appears that in the ditransitive construction, the occurrence frequency in a genre decreases with the degree of formality of the genre (academic > news > fiction > conversation). However, for the caused-motion construction, the trend is reversed: the frequency of usage increases with genre formality. Contextual analysis may then be used to investigate associated concrete expressions specific to each discipline (Figure 4). Such statistical and visual comparisons may help cultivate students' awareness of disciplinary variations of linguistic structures and help them formulate in their writing types of constructions most appropriate for the target discipline in ESP learning.

[3] http://corpus.byu.edu/coca
[4] https://nlp.stanford.edu/nlp/javadoc/javanlp/edu/stanford/nlp/semgraph/semgrex/SemgrexPattern.html

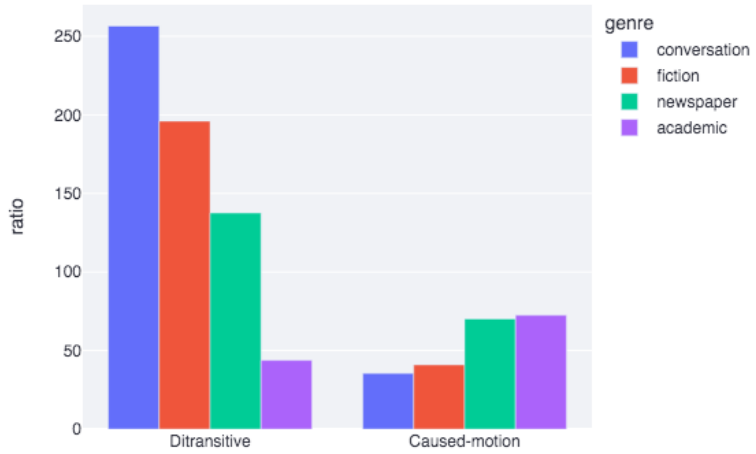

Fig. 3. Frequency ratios (per 100,000 tokens) of two ASCs across genres displayed on the system

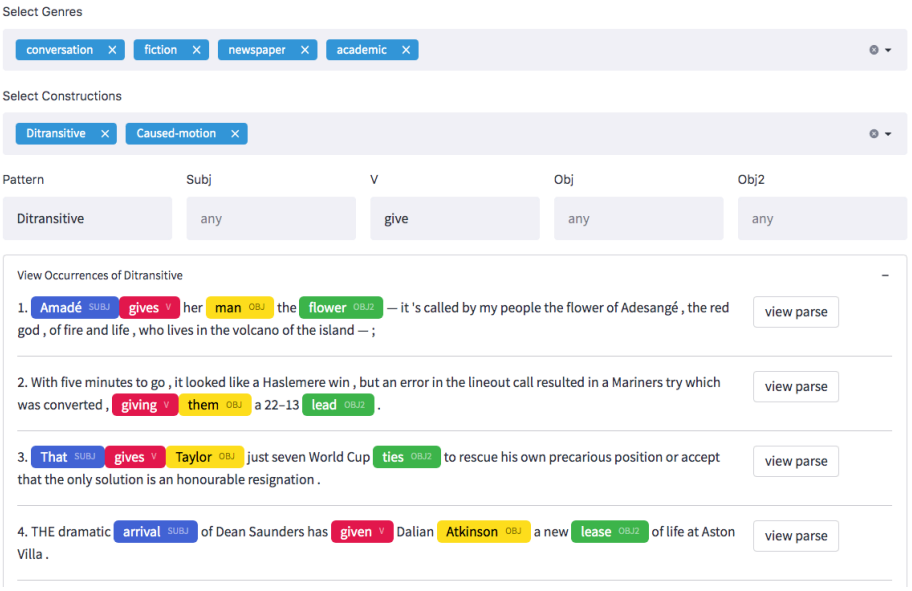

Fig. 4. A view of specific sentences with target constructions in selected genres

3.4 Deploying the system to a blended learning environment

With the development of our system, a blended model of data-driven ESP pedagogy integrating classroom-based, teacher-led instruction and student-centered, self-regulated learning can be established. The deployment of the system to a blended learning environment may help overcome limits of teacher-centered classroom learning and enable online data-driven ESP learning, where students, guided by teachers,

actively explore intricate cross-disciplinary patterning in authentic language use. The web-based architecture allows the system to be used on virtually any computer with a browser, including mobile devices, without the temporal and geospatial obstacles that constrain traditional classroom learning. Teachers, serving as task organizers and guides, can define learning targets (e.g. linguistic constructions of disciplinary significance) and assign materials tailored to the progress and needs of students for study and exploration. Students can then embark on a process of contextualized knowledge discovery, raising research questions, formulating hypotheses, querying data, probing retrieved results, before arriving at their own conclusions which they then summarize and submit to teachers for review.

The system enables the investigations of ESP genres from different angles. The real-time search functions allow users to explore large-scale language corpora across different genres. Through aggregated statistics, students get a "whole picture" of the genre, how it compares with other genres and what typical constructions make it stand out from others. Closer inspections of the construction use in authentic contexts with highlighted concordance can provide students with sufficient/abundant language materials for students to absorb, and raise students' awareness of form known to be beneficial to language learning [39]. Such abilities to search for complex, multifaceted constructions allow students to focus on not only the surface form, but also latent dimensions such as semantics and discourse, grasping the subtle differences among constructions in multiple dimensions of information construal.

4 Conclusion

In this paper, we have presented a new data-driven system for blended ESP pedagogy. Built on state-of-the-art NLP components, it features a set of functionalities specifically designed for data-driven ESP learning through an easy-to-use web-based interface. The system provides the central component for creating a blended environment where human expertise and technological automation combine to augment teaching and learning and help cultivate self-regulated, autonomous learners.

Compared with existing tools, the system brings about improvements in speed, flexibility and accuracy in the retrieval of important multifaceted constructions and enables efficient cross-genre investigations crucial for ESP learning. The system allows users to query, retrieve and explore a wide range of constructions, visually inspect their statistical distributions across different genres, and dive into textual details to develop genre awareness by absorbing essential features of target genres. Compatible with a variety of textual data, whether they be system-provided or user-uploaded, the system can process large-scale corpora and allow for the querying of customizable layers of annotation. The system enables self-regulated or teacher-led cross-genre investigation with the free and guided exploration modes, which allow teachers and students of varying linguistic proficiency and technical expertise to employ different functions of the system for specialized tasks.

One limitation of the current study is that, the proposed system, despite its demonstrated potential, has not seen its pedagogical benefits empirically validated in a real-

world blended learning environment. Computer-assisted learning software, however well designed, can only serve its intended pedagogical purpose as an integral part of a dynamic learning environment involving students, teachers and course designers, whose diverse needs must be carefully surveyed and met. While the system has been developed to address perceived shortcomings in existing DDL software tools, the implemented functionalities still need to be fine-tuned, with potential new features added over time, in response to user feedback in actual usage. In view of such needs, one natural future direction of our research is to empirically validate the efficacy and user perceptions of a pedagogical model that incorporates the system in a blended, data-driven ESP learning environment. Once successfully deployed, it has the potential to extend the affordances of data-driven ESP learning to a much broader user base.

5 Acknowledgments

This study was supported by the Humanities and Social Sciences Foundation of the Ministry of Education of China (Grant No. 21YJC740068), Guangdong Planning Office of Philosophy and Social Science (Grant No. GD20CWY15), and the Bilingual Cognition and Development Lab, Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies (Grant No. BCD20202).

6 References

[1] D. Kaufman, "Constructivist issues in language learning and teaching," *Annu. Rev. Appl. Linguist.*, vol. 24, Mar. 2004. https://doi.org/10.1017/S0267190504000121

[2] U. Felix, "E-learning pedagogy in the third millennium: the need for combining social and cognitive constructivist approaches," *ReCALL*, vol. 17, no. 1, pp. 85–100, May 2005. https://doi.org/10.1017/S0958344005000716

[3] V. Simina and M.-J. Hamel, "CASLA through a social constructivist perspective: WebQuest in project-driven language learning," *ReCALL*, vol. 17, no. 2, pp. 217–228, Nov. 2005. https://doi.org/10.1017/S0958344005000522

[4] O. Tarnopolsky, Constructivist blended learning approach: to teaching English for specific purposes. London: Versita, 2012. https://doi.org/10.2478/9788376560014

[5] E. Corino and C. Onesti, "Data-Driven Learning: A Scaffolding Methodology for CLIL and LSP Teaching and Learning," *Front. Educ.*, vol. 4, p. 7, Feb. 2019. https://doi.org/10.3389/feduc.2019.00007

[6] P. Crosthwaite, "Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing," *Aust. Rev. Appl. Linguist.*, vol. 43, no. 2, pp. 169–195, Jul. 2020. https://doi.org/10.1075/aral.00031.cro

[7] T. Hutchinson and A. Waters, *English for specific purposes: A learning-centered approach*. Cambridge, England: Cambridge University Press, 1987.

[8] J. A. Goldsmith, J. Riggle, D. Schiffrin, D. Tannen, and H. E. Hamilton, *The Handbook of English for Specific Purposes*. 2013.

[9] D. D. Belcher, "English for Specific Purposes: Teaching to Perceived Needs and Imagined Futures in Worlds of Work, Study, and Everyday Life," *TESOL Q.*, vol. 40, no. 1, p. 133, 2006. https://doi.org/10.2307/40264514

[10] D. Belcher, "What ESP Is and Can Be: An Introduction," in *English for Specific Purposes in Theory and Practice*, D. Belcher, Ed. 2009, pp. 1–20. https://doi.org/10.3998/mpub.770237
[11] J. Swales, *Genre analysis: English in academic and research settings*, 1. publ., 14. print. Cambridge: Cambridge University Press, 1990.
[12] K. Hyland, "English for Specific Purposes: Some Influences and Impacts," in *Second Handbook of English Language Teaching*, X. Gao, Ed. Springer, Cham, 2019, pp. 337–353. https://doi.org/10.1007/978-3-030-02899-2_19
[13] H. Basturkmen, "Languages for Specific Purposes Curriculum Creation and Implementation in Australasia and Europe," *Mod. Lang. J.*, vol. 96, pp. 59–70, Jan. 2012, doi: 10.1111/j.1540-4781.2012.01297.x.
[14] C. Uber Grosse and G. M. Voght, "The Continuing Evolution of Languages for Specific Purposes," *Mod. Lang. J.*, vol. 96, pp. 190–202, Jan. 2012. https://doi.org/10.1111/j.1540-4781.2012.01297.x
[15] H. Baştürkmen and A. Bocanegra-Valle, "Materials Design Processes, Beliefs and Practices of Experienced ESP Teachers in University Settings in Spain," in *Key Issues in English for Specific Purposes in Higher Education*, vol. 11, Y. Kırkgöz and K. Dikilitaş, Eds. Cham: Springer International Publishing, 2018, pp. 13–27. https://doi.org/10.1007/978-3-319-70214-8_2
[16] H. Basturkmen, "Is ESP a materials and teaching-led movement?," *Lang. Teach.*, vol. 54, no. 4, pp. 491–501, Oct. 2021. https://doi.org/10.1017/S0261444820000300
[17] T. Johns, "Should you be persuaded – two examples of data-driven learning materials," *Engl. Lang. Res. J.*, vol. 4, pp. 1–16, 1991.
[18] M. Chen and J. Flowerdew, "A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom," *Int. J. Corpus Linguist.*, vol. 23, no. 3, pp. 335–369, 2018. https://doi.org/10.1075/ijcl.16130.che
[19] K. Ackerley, "Exploiting a genre-specific corpus in ESP writing: Students' preferences and strategies," in *Corpora in ESP/EAP Writing Instruction: Preparation, Exploitation, Analysis*, 1st ed., M. Charles and A. Frankenberg-Garcia, Eds. New York, NY: Routledge, 2021, pp. 78–99. https://doi.org/10.4324/9781003001966
[20] A. Boulton and T. Cobb, "Corpus Use in Language Learning: A Meta-Analysis," *Lang. Learn.*, vol. 67, no. 2, pp. 348–393, 2017. https://doi.org/10.1111/lang.12224
[21] R. Godwin-Jones, "Data-Informed Language Learning," *Lang. Learn. Technol.*, vol. 21, no. 3, pp. 9–27, 2017.
[22] J. Dong and X. Lu, "Promoting discipline-specific genre competence with corpus-based genre analysis activities," *Engl. Specif. Purp.*, vol. 58, pp. 138–154, 2020. https://doi.org/10.1016/j.esp.2020.01.005
[23] P. Crosthwaite, "Retesting the limits of data-driven learning: feedback and error correction," *Comput. Assist. Lang. Learn.*, vol. 30, no. 6, pp. 447–473, Aug. 2017. https://doi.org/10.1080/09588221.2017.1312462
[24] E. Cotos, S. Link, and S. Huffman, "Effects of DDL Technology on Genre Learning," *Lang. Learn. Technol.*, vol. 21, no. 3, pp. 104–130, 2017.
[25] L. Anthony, "AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom," in *Professional Communication Conference, 2005. IPCC 2005. Proceedings. International*, 2005, pp. 729–737.
[26] M. Scott, "WordSmith tools version 5," *Liverp. Lex. Anal. Softw.*, vol. 122, 2008.
[27] M. H. Lin and J. Y. Lee, "Data-driven learning: Changing the teaching of grammar in EFL classes," *ELT J.*, vol. 69, no. 3, pp. 264–274, 2015. https://doi.org/10.1093/elt/ccv010
[28] A. E. Goldberg, *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press, 1995.

[29] L. Loenheim, B. Lyngfelt, J. Olofsson, J. Prentice, and S. Tingsell, "Constructicography meets (second) language education: On constructions in teaching aids and the usefulness of a Swedish constructicon," in *Applied Construction Grammar*, S. De Knop and G. Gilquin, Eds. De Gruyter, 2016, pp. 327–356. https://doi.org/10.1515/9783110458268-013
[30] C. R. Graham, "Blended learning systems: definition, current trends, and future directions," in *The handbook of blended learning: global perspectives, local designs*, C. J. Bonk and C. R. Graham, Eds. San Francisco, CA: Pfeiffer, 2006, pp. 3–21.
[31] M. A. Valenzuela-Escárcega, G. Hahn-Powell, and D. Bell, "Odinson: a fast rule-based information extraction framework," in *Proceedings of the 12th Language Resources and Evaluation Conference*, Marseille, France, 2020, pp. 2183–2191.
[32] A. E. Goldberg, *Explain me this: creativity, competition and the partial productivity of constructions*. Princeton: Princeton University Press, 2019. https://doi.org/10.1515/9780691183954
[33] H. Kim and Y. Rah, "Effects of Verb Semantics and Proficiency in Second Language Use of Constructional Knowledge," *Mod. Lang. J.*, vol. 100, no. 3, pp. 716–731, Sep. 2016. https://doi.org/10.1111/modl.12345
[34] A. Stefanowitsch and S. T. Gries, "Collostructions: Investigating the interaction of words and constructions," *Int. J. Corpus Linguist.*, vol. 8, no. 2, pp. 209–243, 2003. https://doi.org/10.1075/ijcl.8.2.03ste
[35] J. Year and P. Gordon, "Korean Speakers' Acquisition of the English Ditransitive Construction: The Role of Verb Prototype, Input Distribution, and Frequency," *Mod. Lang. J.*, vol. 93, no. 3, pp. 399–417, Sep. 2009. https://doi.org/10.1111/j.1540-4781.2009.00898.x
[36] BNC Consortium, *The British National Corpus, version 3 (BNC XML Edition)*. 2007. Accessed: Feb. 16, 2018. [Online]. Available: http://www.natcorp.ox.ac.uk
[37] H. Kim and Y. Rah, "Applying Constructionist Approaches to Teaching English Argument Structure Constructions to EFL Learners," *TESOL Q.*, vol. 55, no. 2, pp. 568–592, 2021. https://doi.org/10.1002/tesq.3002
[38] S. Th. Gries and A. Stefanowitsch, "Extending collostructional analysis: A corpus-based perspective on `alternations'," *Int. J. Corpus Linguist.*, vol. 9, no. 1, pp. 97–129, 2004. https://doi.org/10.1075/ijcl.9.1.06gri
[39] L. Ranta and R. Lyster, "Form-focused instruction," in *The Routledge handbook of language awareness*, P. Garrett and J. M. Cots, Eds. New York: Routledge, 2017, pp. 40–56. https://doi.org/10.4324/9781315676494-3

7 Author

Hengbin Yan is an Associate Professor of Applied Linguistics at Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou City, Guangdong Province, China. He obtained his PhD degree in computational and functional linguistics from the City University of Hong Kong, where he also worked as a post-doctoral fellow. He was a visiting scholar at University of California at Berkeley and Lancaster University. His research interests include Computational Linguistics, Corpus Linguistics and Computer-Assisted Language Learning. Most recently, his work has focused on the computational modeling of construction grammar in contexts of second language acquisition.

Article submitted 2022-01-21. Resubmitted 2022-03-14. Final acceptance 2022-04-08. Final version published as submitted by the author.