

A Blended Grammar Learning System Featuring Unsupervised Pattern Discovery

<https://doi.org/10.3991/ijet.v16i16.21857>

Hengbin Yan, Yinghui Li^(✉)

Bilingual Cognition and Development Lab, Center for Linguistics and Applied Linguistics,
Guangdong University of Foreign Studies, China
liyinhui@gdufs.edu.cn

Abstract—Recent developments in cognitive and psycholinguistic research postulate that language learning is essentially the learning of grammatical constructions. An important type of grammatical construction with wide-ranging pedagogical implications is grammar patterns as laid out in Pattern Grammar. While grammar patterns have seen increasing adoption in language pedagogy, existing applications typically follow a paper-based, teacher-centered approach to instruction, which is known to be less effective in grammar learning than blended, learner-centered approaches. In this paper, we propose a blended learning model that integrates web-based technology with classroom-based instruction to facilitate efficient, personalized grammar learning. We present the design and implementation of a blended grammar learning system that provides customizable learning materials for individual learners by discovering important grammar patterns from corpora in an unsupervised manner. Preliminary evaluation shows that the proposed system achieves an accuracy in pattern discovery comparable to systems that rely on manually precompiled pattern lists and hard-coded rules. With a flexible architecture and an easy-to-use interface, the system can play a key role in the creation of a blended learning environment that can be integrated into a wide range of language learning curricula.

Keywords—blended learning, grammar pedagogy, data-driven learning

1 Introduction

Recent developments in cognitive and psycholinguistic theories have established grammatical constructions, or learned pairings of form and meaning, as the central components of language learning [1]. It is argued that language learning is, in essence, the acquisition of grammatical constructions. Grammatical constructions can be manifested in different forms structurally, and one structural representation scheme that has seen increasing pedagogical applications in recent years is grammar patterns. Grammar patterns, proposed by Hunston and Francis as the basic building blocks in Pattern Grammar [2], [3], are a descriptive representation of grammatical patterns using sequences of abbreviated notations of lexical, phrasal and clausal elements [4]. For example, the pattern *V n to-inf* denotes a verb followed by a noun and a to-

infinitive, and can represent examples such as *ask somebody to do something*. This representation scheme presents a transparent and flexible description of structural patterning in natural language, rendering it particularly suited to language teaching [4]. Grammar patterns are ideal for form-focused instruction, where conscious and focused attention to target structures has been found to result in substantial and durable target-oriented gains compared with implicit instruction, especially for second language learners [1]. Existing pedagogical applications of pattern grammar include development of thesaurus-like resources containing teachable constructions and deriving materials for language teaching [5].

1.1 Blended grammar learning

As a “fundamental linguistic resource to communication” [6], grammar is an essential component of foreign language proficiency. However, it is a difficult subject to master, especially for foreign language learners [7]. It also poses significant challenges to language teaching, where teachers are often unable to devote sufficient time and attention to addressing the needs of individual students, who vastly outnumber them. Existing approaches typically follow a teacher-centered approach to grammar instruction, which has been found to be less effective than learner-centered approaches in the learning of grammatical constructions including multiword units [8].

To address these difficulties, recent works have utilized Blended Learning (BL) to engage students in technology-enabled distributed learning. BL is a hybrid approach to learning that combines traditional face-to-face instruction with computer-assisted learning [9]. Studies [9]–[13] suggest that BL helps create a self-regulated, personalized learning environment to facilitate learner independence and autonomy, which has been found to boost student motivation and overall learning effectiveness, especially in language learning, where personalized learning plays an important role. In grammar pedagogy, a number of studies show that the adoption of BL as a pedagogical model can promote learning and teaching efficiency, motivate students for continued participatory learning, while being positively perceived by its users [14]–[16].

Despite their demonstrated benefits, current approaches to blended grammar learning have suffered from a number of shortcomings. First, as the learning materials are not adapted to learners’ varying levels of proficiency and learning styles (e.g. in the case of [15]), students are often forced to passively memorize grammatical knowledge spoon-fed to them through predefined, one-size-fits-all schedules, and are denied the possibility to autonomously study the materials they find interesting. Second, learners receive little immediate feedback during grammar learning. While the importance of immediate feedback in effective grammar learning has been recognized [17], existing blended approaches to grammar pedagogy have not incorporated mechanisms for providing timely personalized feedback to learners, but instead only utilizes technology as a means to facilitate communication among users. For instance, in the studies reviewed, blended learning is deployed mainly as a means of distributing materials through a computer-mediated interface, e.g. via a third-party online tool such as Moodle or Blackboard [14], [16], which, while still useful, may not have utilized the full benefits of the BL model. Third, existing approaches presume the existence of teach-

ing materials that need to be manually compiled and are often expensive to create and difficult to update (see for example, [14]). The compiled materials often present carefully manipulated examples designed by teachers (who are often non-native speakers themselves), and may not be properly adapted to the great variety and complexity of real-world language use as well as the personalized needs of the learners.

Addressing these problems calls for the design and implementation of a blended system that not only serves as a medium for online communication, but also (1) delivers targeted, authentic materials that tightly integrate with the grammar learning process; and (2) adapts to personalized needs by enabling users to engage in data-driven, exploratory learning [18], [19].

1.2 Grammar pattern discovery

To create a blended, personalized learning grammar system, one essential functionality is the automatic discovery of grammar patterns from target texts so as to generate learning resources tailored to the learning purpose at hand. This task is known as pattern discovery [20], which involves finding new grammar patterns that are unknown in advance. The process takes textual corpora as input, and outputs candidate patterns as a repository of patterns in an unsupervised manner for downstream tasks. Pattern discovery has not been implemented in previous systems for automated grammar pattern analysis [21]–[23], which have instead focused exclusively on pattern identification [20], a task concerned with annotating texts with occurrences of known grammar patterns using a precompiled list of patterns and supervised/rule-based methods. These existing studies share some similar shortcomings: (1) they are supervised, i.e. relying on the pre-existence of a manually compiled pattern list, which can be time-consuming and expensive to build and update; and (2) they require manually hard-coded rules to extract each individual pattern programmatically. For building personalized learning systems, pattern identification is inadequate since the manually compiled, “one-size-fits-all” pattern lists they rely on do not reflect the constantly changing patterns in use and do not adapt to context- and domain-specific requirements in a blended learning environment. For many specific genres, where a specialized pattern list is not readily available, an unsupervised pattern discovery algorithm is desirable for replacing or at least speeding up the traditionally expensive process of pattern compilation.

1.3 The proposed blended grammar learning system

We propose in this paper a blended grammar learning system featuring unsupervised pattern discovery from corpora. The system automates the discovery of grammar patterns using a simple and efficient algorithm applicable to large-scale language data, providing an extendable architecture and a user-friendly interface that can be easily integrated into a blended language learning curriculum. In the remainder of the paper, we describe the process of developing the system using a web-based architecture, evaluate the accuracy of grammar pattern discovery, and discuss ways to integrate the system into a blended grammar learning curriculum.

2 Method

2.1 System architecture

Our proposed system adopts a web-based architecture that consists of several interconnected modules, forming a workflow as illustrated in Figure 1.

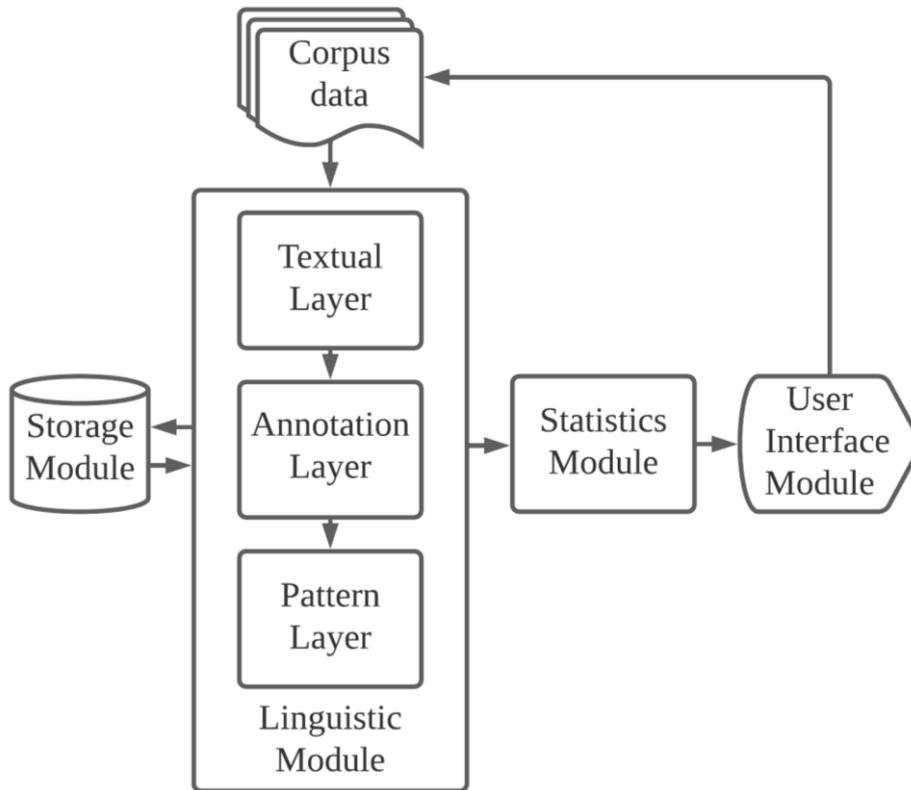


Fig. 1. Overall architecture of the proposed system

The functionalities provided by the modules are as follows:

1. The storage module is responsible for storing the different layers of linguistic information in a relational database. The linguistic module produces several layers of information, each handled by an individual processor:
 - a. The textual layer indexes text in various formats (currently compatible formats include raw texts and TEI-compliant XML formats) uploaded by users. This generic representation allows arbitrary layers of linguistic annotation (e.g., lexical, syntactic, and semantic) to be aligned and analyzed, making it flexible for the analysis of various types of linguistic information.

- b. The annotation layer preprocesses each stored text using state-of-the-art syntactic parsers for lexical and syntactic annotations.
 - c. The pattern layer employs an innovative algorithm to process the previously stored syntactic information for grammar pattern extraction.
2. The statistics module collects and aggregates relevant statistics (e.g. token pattern frequencies) and sends them to the user interface module.
 3. The user-interface module features a web-based interface that allows users to perform corpus searches, examine grammar patterns, and explore general linguistic patterning through tabular data and visual graphs.

The frontend browser-based interface of the system was built using the JavaScript Programming language along with the Vue.js web framework, while the backend server employed the Django web framework. System modules were implemented with the Python programming language (version 3.8) to enable efficient communication between different Natural Language Processing (NLP) and corpus-processing components (which include third-party libraries also implemented in Python).

2.2 Corpus data

The proposed system can operate on arbitrary textual data for pattern discovery and analysis. With its efficient storage and processing pipeline, it can work with large-scale textual corpora. The 100-million-word British National Corpus (BNC) XML edition is used as the default corpus for pattern discovery due to its balance and comprehensiveness in representing British English, which is the main dialect taught around the world as a second or foreign language. For the evaluation of pattern discovery, we use a balanced subset of the BNC, the BNC baby, to make intensive computation and manual evaluation feasible. The BNC baby is divided into four equally sized sub-corpora totaling four million words, each representing text/speech from a distinct genre: news, academic, fiction and conversation.

2.3 Algorithm for pattern discovery

Our algorithm for pattern discovery is inspired by key insights derived from recent advances in corpus and cognitive linguistics. Corpus-based studies have long established various association measures (AMs) to quantify the degree of association between collocations, pairs of tokens that co-occur with higher-than-chance probabilities [24]. Commonly used AMs include Pointwise Mutual Information, t-score, and Log-Likelihood Ratio etc. Such investigations are often limited to bigrams, since there is no easy way to extend the pairwise AMs to multiword units. One simple and effective strategy is to merge bigrams recursively to form multiword sequences. The strategy, however, has not been applied to grammar pattern discovery. Another problem with existing studies is their confinement to the lexical level – specific sequences of lexical tokens, which fails to capture generalizable schemas of expressions essential for making language productive and expressive. We adopt a more generic algo-

rithm that extends existing association measures to multi-unit, multilayered sequences for the investigation of mid-level constructions like grammar patterns.

Another source of algorithmic inspiration comes from cognitive construction grammar [25]. The algorithm simulates the gradual learning of constructions by humans [26], pairing highly associated tokens incrementally to form increasingly large chunks and units. Similar to human learning [26], the textual inputs are chunked at multiple levels, from words up to phrasal/multiunit sequences. The order of chunking is determined by frequency and saliency operationalized as ngram frequency and AMs. As in human learning [1], [26], the sequences of utterances do not always have clear-cut boundaries. One advantage of this approach is that overlapping sequences can be automatically dealt with, since enclosing and enclosed sequences are cleanly separated. This is especially useful in subsequent statistical analysis, where overlapping sequences can be associated with different meanings/functions (e.g. the meaning of *put up* can be different from *put up with*, *put up for*, *put up to*, *put up against* etc.)

We analyze the syntactic structure of each text using the Stanford parser, an efficient parser for corpus analysis with state-of-the-art performance [27]. The parser first splits each text into separate sentences and tokens. Each token in each sentence is then assigned three levels of structured annotation: lemmas, part-of-speech tags (based on the Penn treebank tag set) and a constituency parse tree. The parse tree is a hierarchy consisting of nodes of clauses, phrases and tokens. Clausal and phrasal structures are represented by spanning nodes in the tree. Leaf nodes (tokens) in the tree may be covered by multiple spanning clausal and phrasal labels simultaneously. For example, for the verbal phrase *go to the park*, *the park* is part of a Noun Phrase (NP), and Prepositional Phrase (PP) and Verb Phrase (VP) in the parse tree at the same time. For ease of subsequent processing, we convert each parse tree into a layered structure, where one layer represents a label spanning one or more words. To search for patterns, the algorithm first scans for bigrams formed by combining words and phrases within a certain window size across different layers. Each bigram is a candidate pattern or component for a larger candidate pattern (depending on whether it can merge with more components in subsequent steps). Each token in the text is assigned one or more tags corresponding to its layers, and each tag will form a bigram with each of its adjacent (or non-adjacent if the window-size is greater than 1) tags. This process is repeated for each word in the text, with the frequencies of unigram and bigram occurrences counted.

Table 1. Example layered structure for token, part-of-speech (POS) and constituency parse labels

1	Token	Mary	loves	playing	basketball
2	POS	NNP	VBZ	VBG	NN
3	Constituency	NP			NP
4				VP	
5				S	
6			VP		
7		S			

To search for patterns, the algorithm first scans for bigrams formed by combining words and phrases within a certain window size across different layers. Each bigram is a candidate pattern or component for a larger candidate pattern (depending on whether it can merge with more components in subsequent steps). Each token in the text is assigned one or more tags corresponding to its layers, and each tag will form a bigram with each of its adjacent (or non-adjacent if the window-size is greater than 1) tags. This process is repeated for each word in the text, with the frequencies of unigram and bigram occurrences counted.

After all the bigram combinations have been exhausted, the collected frequencies are used to compute the strength of each bigram association using the chosen association measure. We choose the Log-Likelihood Ratio due to its balanced performance [24] and its capacity to select both schematized/general and specific collocations. The discovery process iterates over a number of rounds. In each round, the top candidate (bigram with the highest association score) is extracted. The bigram is then combined into a single unit, and all the text in the corpus matching the bigram is identified and replaced with this new form. In the subsequent round, all the frequencies and association scores for all bigrams are recomputed. The newly merged form will serve as a single unit that can be combined again with other units to form larger units. This process continues until a designated number (e.g. 20) of rounds is reached (when a designated number of constructions have been discovered) or none of the remaining association scores is above a statistically significant threshold (for the recommended critical values for the log-likelihood ratio, see [28]).

3 Results and discussion

3.1 Web-based frontend

Based on the architecture proposed in Section 2, we implemented an online blended learning system adopting a client-server model. As shown in Figure 2., the browser-based frontend features a dashboard-like interface allowing for easy user access and customization. The interface consists of several interactive components for corpus management, automated text annotation and statistics/visualization. Each of the frontend components corresponds to a module in the backend architecture: The corpus management component corresponds to the storage module, the automated text annotation component to the linguistic module, while the statistics and visualization component, which generates and updates statistical tables and visual graphs in real time in response to user queries, is connected with the statistics module.

Typical scenarios for blended grammar learning utilizing the system involve the following three steps:

1. Instructors upload course-related corpus data to the system and create individualized learning sessions, performing automated mining of grammar patterns on the data.

2. Students engage in exploratory studies by performing pattern searches in instructor-assigned or self-uploaded materials, investigating statistical and visual summaries of the patterns. The pattern distribution statistics provide an overview of the frequency information of each pattern, revealing its distribution across different subdivisions (e.g., genres).
3. Teachers monitor and track the learning progress of student accounts on the system backend so as to provide students with targeted, personalized feedback.

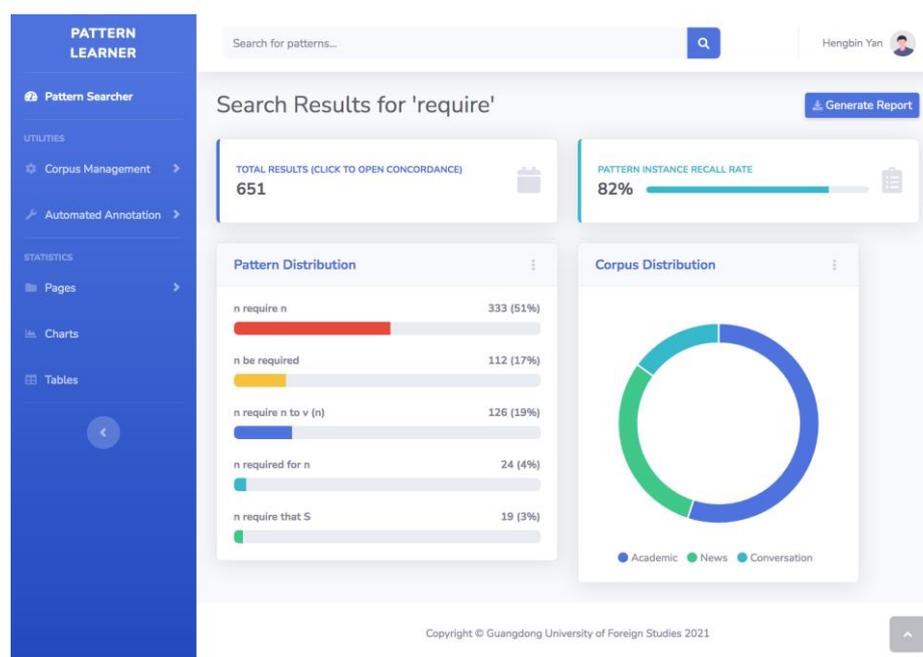


Fig. 2. A screenshot of the web-based user interface

3.2 Evaluation on grammar pattern discovery

We evaluate the performance of the pattern discovery of the system through comparison with two manually produced lists: First, a list of patterns manually extracted from the same corpus that the program runs on; second, the verb pattern list in [21], consisting of patterns for 115 verbs taken from the Collins COBUILD English Dictionary and manually verified using a similarly sized corpus. The academic and news portions of the BNC Baby corpus with a size of about 1 million words each are used as data on which pattern discovery is performed. Two language experts independently examined the concordance context of the target word to annotate them as patterns using elements of pattern grammar (which can include the target word, prepositions, verbs, nouns, NPs, VPs etc.). Patterns attested by examples whose frequencies are lower than a particular threshold (in our case 3) are filtered out. The annotation agreement between the two experts is then calculated. The remaining differences are

resolved through a thorough discussion between the annotators. The resulting list represents a gold standard against which the pattern discovery algorithm can be evaluated, by computing the degree to which the discovered patterns matched the gold standards. As different tagging schemes for defining grammatical patterns are used between constituency parsing and pattern grammar, we adopt a number of mappings for conversion. For example, the original output from the syntactic parser NP is converted to the tag “n” (meaning noun). In addition, as the automatically induced boundaries are not always the same as the manually truncated list, we stipulate that as long as one is contained in the other, as judged by the annotator, it will be considered a match. A comparison between the gold standard list and the automatically discovered list can result in three possibilities: (1) the patterns are present in both the automatic discovery and the gold standard list; (2) valid, newly discovered patterns are not present in the gold standard list; (3) patterns are present in the gold standard list but missing in the automatically discovered list (potentially due to the lack of sufficient examples, or errors in the algorithm).

The precision of the pattern discovery results is 0.844, with a recall rate of 0.865. The overall accuracy in terms of the F1 score, computed using the following formula, is 0.854.

$$F1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (1)$$

The score is comparable to Ma and Qian’s score of 0.881 (computed using the provided precision and recall) [21]. Given that Ma and Qian’s results were obtained using a supervised approach, relying on the existence of a precompiled pattern list and programmatically defined procedures customized for the retrieval of each pattern, the unsupervised nature of our proposed algorithm may yield benefits where such a list or programmatic resources are not readily available or cost-effective. Table 2 shows a statistical summary of grammar patterns of the five most frequent verbs from the full verb pattern list.

The results of pattern discovery show that, by following a process inspired by corpus linguistics measures and a cognitively motivated language acquisition model, the automatically discovered grammar patterns of common verbs are roughly comparable to expert-crafted patterns. In some cases, the patterns are not as comprehensive as the expert patterns, presumably due in part to the relatively small corpus size preventing the algorithm from generalizing a few examples to a pattern. In other cases (roughly 40% of all patterns), however, the patterns have produced a richer summary that exceeds the suggested pattern list for a given verb, presumably attributable to the richer, more fine-grained contexts in which the patterns appear in different genres. One major difference between the automatically discovered patterns and expert patterns is that the discovered patterns do not always have a clear-cut boundary. This is due to the algorithm’s effort to reflect the nature of constructional learning: the core/top-ranked components of a construction might carry more prototypical meaning, while the other components radiate from the prototype along a cline of sequential unity.

Table 2. Statistical summary of grammar patterns of the five most frequent verbs.

Target verb	Discovered patterns	Count	Precision	Recall	F1 Score
include	n include n be included include n in n be included	973 16 78 24	0.83	0.96	0.89
provide	provide n provide n for n n be provided by n provide n with n provide n to n be provided provide that S	503 229 89 65 45 56 27	0.85	0.93	0.89
involve	involve in doing involve in n n involve n be involved Involve with	36 161 446 48 8	0.53	0.95	0.68
suggest	n suggest that S n suggest n as n suggest n suggested by n suggest n be n be suggested	451 232 41 15 12 38	0.94	0.91	0.92
describe	be described n describe n n be described n describe n as n be described in n be described by as described	21 345 31 172 79 16 32	0.90	0.91	0.90

3.3 Error analysis

Despite the relatively high accuracy, a number of recurring errors appear to reveal the shortcomings of the completely unsupervised algorithm. One prominent problem is the inclusion of high-frequency elements that are not an integral part of the pattern. For example, around 17% of the samples for the verb "include" contain the subsequence "NP of NP", which, while a valid and common nominal pattern by itself, is not an integral part of the verbal pattern and is therefore considered incorrect, significantly reducing the overall accuracy for the verb. While such errors can be trivially corrected using rules, new heuristics may be needed to arrive at a generalized solution to similar cases.

3.4 Integration into a grammar teaching curriculum

The proposed blended grammar learning system is designed to be easily integrated into a grammar teaching curriculum. Apart from its obvious use as a thesaurus-like resource for students and teachers to look up grammar patterns, the system can be used as a self-regulated learning tool to:

1. Discover grammar patterns in corpora of interest. The corpora can be of any genre and consist of learning materials tailored to a particular learning purpose, e.g. an ESP (English for Specific Purpose) course in accountancy. The results are returned in the form of a list of patterns mined from the corpora. With its flexible architecture, the tool can be applied to the discovery of grammatical constructions at varying degrees of schematicity.
2. Discover pattern usage in texts. The tool can be used to discover patterns in any given text. The text can be sourced from a textbook lesson to raise awareness on important usage in the text. It can also be a text written by the learner, to be used to provide personalized feedback for the learner/teacher. The interface allows the user to click on a pattern to view contextualized sample occurrences and freely explore the inventory of patterns for a particular genre, or a particular text.
3. View descriptive statistics and visualizations. Statistics including the type and count of each pattern, each component in the pattern, and associated contexts, and genres etc. are collected, summarized and visualized on a web interface. This process is highly efficient since all patterns have been indexed during pattern discovery.

The system thus makes possible a blended learning environment that can be used to engage students in both teacher-guided and self-regulated learning in and outside of the classroom. In class, the system can help teachers design teaching materials throughout the curriculum. Traditionally grammar is taught in dedicated grammar courses, but the system can be blended with many language learning courses for teaching the language skills required in specific domains. Teachers use the system to perform pattern discovery on reading texts and automatically retrieve grammar patterns linked with authentic example usage. The retrieved example patterns from the course materials can be corroborated by or contrasted with those mined from other corpora, potentially from another genre to demonstrate register differences. Teachers can draw students' attention to various usage statistics of similar patterns across different genres and contexts to heighten their awareness of contextual appropriateness when using those patterns. Thus, the system becomes an ideal tool for studying language for academic and other specific purposes. After class, exercises can be assigned to students for them to explore the subtle usage of various target words and associated patterns. For example, blank filling exercises with provided keys and instant feedback can be easily created on the platform to consolidate students' mastery of target patterns. Students can engage in data-driven learning tailored to their personalized needs and proficiency levels (which can be predefined in the system and adapt as students progress through exercises). Aided by the interactive functions of the system, messages can be sent in the system to the teacher and between peer students when needed. Before new lessons, teachers can view individual and collective learner statistics to better prepare materials and exercises suitable for the current cohort of students. Potential problems with individual students can also be discovered and intervention administered in a timelier manner.

In summary, the blended learning model made possible by the system yields flexibility not afforded by traditional paper-based grammar classrooms, allowing students

to learn according to personalized preferences and proficiencies, unlimited by time, geographic location and teacher-student ratios.

4 Conclusion

In this paper, we proposed a blended grammar learning system featuring unsupervised discovery of grammar patterns. We described the main algorithms for the system, evaluated the accuracy of pattern discovery, and discussed ways in which the system can be integrated into a blended, data-driven learning curriculum, where the proposed technology joins hands with a student-centered pedagogy to cultivate smart, autonomous learners. The system allows for customized analysis of user-provided corpora, enabling self-regulated, personalized learning of language materials of varying proficiency levels, genres and interests. Adopting the blended learning model, the system can: (1) help teachers design teaching materials (e.g. thesaurus-like resources and exercises); (2) help students personalize their grammar learning; and (3) serve to mediate between teachers and students in a data-driven learning environment. As a work in progress, however, the system still requires further development and additional empirical validation for it to function as an efficient, full-fledged blended learning platform.

5 Acknowledgments

This study was supported by two grants from Guangdong Planning Office of Philosophy and Social Science (Grant No. GD18YWW02; Grant No. GD20CWY15), and a research grant (No. BCD20202) from the Bilingual Cognition and Development Lab, Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies.

6 References

- [1] S. Wulff and N. C. Ellis, “Usage-based approaches to second language acquisition,” in *Bilingual Cognition and Language. The state of the science across its subfields*, 2018, pp. 37–56. <https://doi.org/10.1075/sibil.54.03wul>
- [2] S. Hunston and G. Francis, *Pattern Grammar*, vol. 4. Amsterdam: John Benjamins Publishing Company, 2000.
- [3] S. Hunston, “Patterns, constructions, and applied linguistics,” *Int. J. Corpus Linguist.*, vol. 24, no. 3, pp. 324–353, 2019. <https://doi.org/10.1075/ijcl.00015.hun>
- [4] S. Hunston, “Pattern grammar, language teaching, and linguistic variation,” in *Using corpora to explore linguistic variation*, R. Reppen, S. M. Fitzmaurice, and D. Biber, Eds. Amsterdam: John Benjamins, 2002, pp. 167–183. <https://doi.org/10.1075/scl.9.11hun>
- [5] S. Hunston and H. Su, “Patterns, constructions, and local grammar: A case study of ‘evaluation,’” *Appl. Linguist.*, vol. 40, no. 4, pp. 567–593, 2019. <https://doi.org/10.1093/applin/amx046>

- [6] J. E. Purpura, “Assessing Grammar,” in *The Companion to Language Assessment*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013, pp. 100–124.
- [7] R. M. DeKeyser, “What Makes Learning Second-Language Grammar Difficult? A Review of Issues,” *Lang. Learn.*, vol. 55, no. S1, pp. 1–25, Jun. 2005. <https://doi.org/10.1111/j.0023-8333.2005.00294.x>
- [8] S. Yamagata, “Comparing core-image-based basic verb learning in an EFL junior high school: Learner-centered and teacher-centered approaches,” *Lang. Teach. Res.*, vol. 22, no. 1, pp. 65–93, 2018. <https://doi.org/10.1177/1362168816659784>
- [9] J. Poon, “Blended learning: an institutional approach for enhancing students’ learning experiences,” *J. Online Learn. Teach.*, vol. 9, no. 2, pp. 271–288, 2013.
- [10] C. R. Graham, W. Woodfield, and J. B. Harrison, “A framework for institutional adoption and implementation of blended learning in higher education,” *Internet High. Educ.*, vol. 18, pp. 4–14, 2013. <https://doi.org/10.1016/j.iheduc.2012.09.003>
- [11] J. Jia, Y. Chen, Z. Ding, and M. Ruan, “Effects of a vocabulary acquisition and assessment system on students’ performance in a blended learning class for English subject,” *Comput. Educ.*, vol. 58, no. 1, pp. 63–76, Jan. 2012. <https://doi.org/10.1016/j.compedu.2011.08.002>
- [12] T. Miyazoe and T. Anderson, “Learning outcomes and students’ perceptions of online writing: Simultaneous implementation of a forum, blog, and wiki in an EFL blended learning setting,” *System*, vol. 38, no. 2, pp. 185–199, 2010. <https://doi.org/10.1016/j.system.2010.03.006>
- [13] Y. T. C. Yang, Y. C. Chuang, L. Y. Li, and S. S. Tseng, “A blended learning environment for individualized English listening and speaking integrating critical thinking,” *Comput. Educ.*, vol. 63, pp. 285–305, 2013. <https://doi.org/10.1016/j.compedu.2012.12.012>
- [14] I. Simonova, “Blended approach to learning and practising English grammar with technical and foreign language university students: comparative study,” *J. Comput. High. Educ.*, vol. 31, no. 2, pp. 249–272, Aug. 2019. <https://doi.org/10.1007/s12528-019-09219-w>
- [15] A. Zibin and A. R. M. Altakhaineh, “The effect of blended learning on the development of clause combining as an aspect of the acquisition of written discourse by Jordanian learners of English as a foreign language,” *J. Comput. Assist. Learn.*, vol. 35, no. 2, pp. 256–267, Apr. 2019. <https://doi.org/10.1111/jcal.12327>
- [16] A. M. Pinto-Llorente, M. C. Sánchez-Gómez, F. J. García-Peñalvo, and S. Casillas-Martín, “Students’ perceptions and attitudes towards asynchronous technological tools in blended-learning training to improve grammatical competence in English as a second language,” *Comput. Human Behav.*, vol. 72, pp. 632–643, Jul. 2017. <https://doi.org/10.1016/j.chb.2016.05.071>
- [17] F. Kılıçkaya, “Pre-service language teachers’ online written corrective feedback preferences and timing of feedback in computer-supported L2 grammar instruction,” *Comput. Assist. Lang. Learn.*, vol. 0, no. 0, pp. 1–26, 2019. <https://doi.org/10.1080/09588221.2019.1668811>
- [18] J. Geluso and A. Yamaguchi, “Discovering formulaic language through data-driven learning: Student attitudes and efficacy,” *ReCALL*, vol. 26, no. 2, pp. 225–242, 2014. <https://doi.org/10.1017/s0958344014000044>
- [19] 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>
- [20] M. Constant *et al.*, “Multiword Expression Processing: A Survey,” *Comput. Linguist.*, vol. 43, no. 4, pp. 837–892, Dec. 2017.

- [21] H. Ma and M. Qian, “The creation and evaluation of a grammar pattern list for the most frequent academic verbs,” *English Specif. Purp.*, vol. 58, pp. 155–169, Apr. 2020. <https://doi.org/10.1016/j.esp.2020.01.002>
- [22] O. Mason and S. Hunston, “The automatic recognition of verb patterns: A feasibility study,” *Int. J. Corpus Linguist.*, vol. 9, no. 2, pp. 253–270, 2004. <https://doi.org/10.1075/ijcl.9.2.05mas>
- [23] U. Römer, M. B. O’Donnell, and N. C. Ellis, “Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions,” in *Corpora, Grammar and Discourse*, M. C. and S. J. Nicholas Groom, Ed. John Benjamins Publishing Company, 2015, pp. 43–71. <https://doi.org/10.1075/scl.73.03rom>
- [24] B. Kang, “Collocation and word association,” *Int. J. Corpus Linguist.*, vol. 23, no. 1, pp. 85–113, 2018.
- [25] A. E. Goldberg, *Constructions at work: the nature of generalization in language*. Oxford: Oxford University Press, 2006.
- [26] M. H. Christiansen and N. Chater, “The Now-or-Never bottleneck: A fundamental constraint on language,” *Behav. Brain Sci.*, vol. 39, pp. 1–72, Apr. 2015. <https://doi.org/10.1017/s0140525x1500031x>
- [27] D. Chen and C. Manning, “A fast and accurate dependency parser using neural networks,” in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 740–750. <https://doi.org/10.3115/v1/d14-1082>
- [28] P. Rayson, D. Berridge, and B. Francis, “Extending the Cochran rule for the comparison of word frequencies between corpora,” in *7th International Conference on Statistical analysis of textual data (JADT 2004)*, 2004, pp. 926–936.

7 Authors

Hengbin Yan is an Associate Professor of Applied Linguistics at Bilingual Cognition and Development Lab, 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.

Yinghui Li is a Lecturer of Psycholinguistics at Bilingual Cognition and Development Lab, Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou City, Guangdong Province, China. She obtained her Ph.D. in Psycholinguistics at Guangdong University of Foreign Studies. Her research interest is bilingual language processing, in particular the cognitive and psychological process of interpreting.

Article submitted 2021-02-15. Resubmitted 2021-04-22. Final acceptance 2021-05-16. Final version published as submitted by the authors.