Applied Linguistics, 2024, 45, 786–805
https://doi.org/10.1093/applin/amae001
Advance access publication 25 January 2024
Article

What Should Go With This Word Here: Connecting Lexical Collocations and Rhetorical Moves in Narrative Stories

Yishi Jiang[1], Xiaofei Lu[2], Fengkai Liu[3], Jianxin Zhang[4], and Tan Jin[5,*]

[1]School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China
[2]Department of Applied Linguistics, The Pennsylvania State University, University Park, PA, USA
[3]Department of Linguistics and Translation, City University of Hong Kong, Hong Kong, China
[4]School of International Studies, University of International Business and Economics, Beijing, China
[5]School of Foreign Studies, South China Normal University, Guangzhou, Guangdong, China

[*]E-mail: tjin@scnu.edu.cn

An emerging body of corpus-based genre analysis studies has examined the connection between different types of formulaic language and rhetorical moves in various genres of academic writing. The current study extends this body of research into the understudied genre of narrative stories and the understudied phraseological unit of lexical collocations. Specifically, we compiled a corpus of narrative stories written by expert writers, extracted a list of frequent collocations from the corpus, developed a rhetorical move framework for narrative stories, examined the distribution of rhetorical stages and moves in the corpus, and explored the connection between collocations and rhetorical moves in the corpus. The findings of our research culminated in an online interface for searching the corpus for collocations and exploring their use in sentences realizing different rhetorical stages and moves in context. We discuss the potential pedagogical value of our findings and the resulting online interface for promoting learner awareness of the connection between linguistic features and rhetorical functions in narrative stories in genre-based pedagogy.

Introduction

An increasing body of research has employed corpus-based text analysis and genre-based move analysis to unveil the linguistic and rhetorical features of academic (e.g. Cortes 2013; Mizumoto *et al.* 2017) and argumentative genres (e.g. Bychkovska and Lee 2017). Such research has generated useful findings regarding the formulaic and functional patterns of discourse, bridging the form-function gap in genre analysis research and yielding pedagogical resources that can help novice writers participate in unfamiliar genres (Tardy 2016). However, little research in this area has attended to narrative stories, an important genre frequently used to help young second language (L2) learners become proficient in interpreting information in the target language and using it to represent their worlds to others (Lawler 2002; Berman and Nir-Sagiv 2007). Familiarity with the linguistic patterns and plot sequences of narratives is essential for learners to produce grammatically well-formed and logically well-developed narratives (Deane *et al.* 2019). A

systematic analysis of the types of formulaic language used to communicate different rhetorical functions in narrative stories can contribute findings to inform genre-based pedagogy aimed at helping learners develop this competence. In light of these considerations, the current study undertakes to identify frequently occurring lexical collocations in a corpus of expert-composed narrative stories, construct a rhetorical move framework for narrative stories, match the identified collocations to rhetorical moves in the corpus, and discuss the implications of our findings for genre-based writing research and pedagogy.

Lexical collocations

Lexical collocations, that is, frequently co-occurring pairs of words that are grammatically connected (Nation 2013), are one type of multi-word expressions that have received much attention in formulaic language research. Knowledge of lexical collocations is essential for L2 learners (Webb and Kagimoto 2011), as evidenced in Nation's (2013) claim that 'language knowledge is collocational knowledge' (517). Indeed, as important building blocks of spoken and written discourse, lexical collocations can represent the distinct characteristics of different genres, making it important for learners to be aware of the usage of different lexical collocations in different genres of text (Conrad and Biber 2004). In terms of language processing and production, experiencing patterns of word co-occurrence in context and storing such patterns in their long-term memory can help learners improve the fluency and thoroughness of their language processing as well as the fluency and formulaicity of their language production (Underwood *et al.* 2004; Ellis 2012).

Lexical collocations are morphologically flexible phraseological items. Different from word combinations presented in a fixed order (e.g. lexical bundles), lexical collocations are built around a *pivot word* and a *collocate* that may occur in different orders and morphological forms, possibly with a certain number of intervening words between them. A pivot word is usually a high-frequency word that is used to search for other words that frequently co-occur with it, namely, its collocates (e.g. Rogers *et al.* 2021). For instance, a search for the lemmatized pivot word *advice* and its collocate *give* in a text may yield *giving advice*, *advice that she gives*, *gave me some advice*, etc., all of which are considered instances of the same lexical collocation. This high flexibility in form makes lexical collocations a highly productive and yet potentially challenging resource for language learners.

In identifying lexical collocations, researchers have usually considered pivot words and collocates that are both content words that enter into a particular grammatical relationship, and that co-occur frequently across multiple texts (e.g. Wray and Perkins 2000; Ackermann and Chen 2013). These criteria help ensure the semantic meaningfulness, grammatical integrity, and a high enough frequency of use of the lexical collocations identified. With these criteria, researchers have extracted different lists of lexical collocations from different types of text corpora. For instance, Shin and Nation (2008) identified the top 1,000 most frequent collocations in spoken English and argued that most of them could be taught in an elementary English course. Durrant (2009) extracted a list of 1,000 positionally variable academic collocations from a corpus of research articles (RAs) and reported that students in different disciplines may have different needs when it comes to learning those collocations. Ackermann and Chen (2013) extracted 2,468 frequent lexical collocations from a corpus of academic English that were judged by a panel of experts to be pedagogically relevant.

Rhetorical moves in narrative stories

The concept of *genre* is based on the idea that members of a discourse community can recognize the similarities among texts frequently used in their community and draw on their experiences to comprehend and produce such texts relatively easily (Hyland 2008). Specifically, a genre is a class of communicative events with a common set of communicative purposes (Swales 1990), and *genre analysis* aims at revealing the rhetorical structures of texts in a particular genre, commonly

involving what is known as rhetorical move analysis (Swales 1981). A rhetorical move is 'a section of text that performs a specific communicative function' (Kanoksilapatham 2007: 23). For example, Swales (1990) identified three rhetorical moves in RA introductions, namely, establishing a territory, establishing a niche, and occupying a niche, along with a set of steps characterizing the organization of each move. Having learners perform genre analysis of texts in a particular genre has the potential of raising learners' awareness of the textual and rhetorical features of that genre (Negretti and McGrath 2018).

Narratives are a literary genre that aims at accounting 'connected events or people's actions in a specific time frame' (Yoon 2021: 113), in which events or affairs are generally chronologically ordered and causally related (Sanford and Emmott 2012). Narratives are considered as an 'essential genre that provides important information about the narrator's linguistic competence and pragmatic sensitivity in the target language' (Kang 2005: 260) and that is crucial to literate life and literate achievement (Deane *et al.* 2019). Through the process of learning and crafting narrative stories, learners can enhance their language skills, gain mastery of the structures of narrative texts, articulate their thoughts, nurture their self-identity, and foster meaningful social relationships (Jović 2020).

The overall structure of a narrative, according to Labov and Waletzky (1967), usually consists of five 'stages'. These include an *orientation* that sets the scene and the context, a *complication* that disrupts the event, an *evaluation* of the protagonist's feelings and judgement, followed by a *resolution* that presents the solutions to problems and an optional *coda* that concludes the sequence of events and returns the reader/audience to the reality of the present. The complication–evaluation–resolution sequence may be recurrent, creating a multiple-event narrative story. Drawing on previous literature, Flynn (2018) formulated and categorized 10 functional stages of a story in order to examine how young children's construction of experience was related to the functions in oral storytelling. Flynn's functional stages combine 'the purposes that the stages serve' and the 'patterns of realization in the grammar of language' (9), filling the form-function gap in oral storytelling.

The connection between lexical collocations and rhetorical moves

While rhetorical moves and formulaic language represent different levels of textual features, the association between the two constructs provides evidence for the integration of organizational and grammatical patterns used to build discourse (Mizumoto *et al.* 2017). Recent research on connecting rhetorical moves and different types of formulaic expressions has underscored the value of analyzing the synergy between organizational patterns and lexico-grammatical features in achieving a holistic understanding of discourse construction across diverse genres (e.g. Lu *et al.* 2021a). For example, Cortes (2013) identified lexical bundles from a corpus of RA introductions, categorized the bundles into four structural types and three functional types using Biber *et al*'.s (2004) taxonomies for structural and functional categorization, and connected bundles to rhetorical moves in RA introductions. Mizumoto *et al.* (2017) connected 250 lexical bundles and 25 rhetorical moves in applied linguistics RAs through a correspondence analysis. Omidian *et al.* (2018) investigated disciplinary variation in the use of multi-word expressions in rhetorical moves of RA abstracts. Incorporating the classification frameworks proposed by Biber *et al.* (2004) and Hyland (2008), they categorized the multi-word expressions into three distinct groups based on their communicative functions in different moves: *research-oriented*, *text-oriented*, and *participant-oriented*. They reported substantial variation in the use of multi-word expressions that fulfill different communicative purposes between RA writers in hard and soft disciplines, reflecting differences in their priorities in reporting their research in RA abstracts. Recently, Lu *et al.* (2021b) matched phrase-frames with rhetorical moves of social science RA introductions and categorized the phrase-frames based on whether they occur only in one move-step, primarily in one move-step, or across multiple move-steps. In a similar analysis, Casal and Kessler (2020)

examined the form and rhetorical function of phrase-frames in a promotional genre (i.e. purpose statements in Fulbright grant applications). They identified 69 phrase-frames from 148 purpose statements, assigned them to one of four rhetorical move codes based on their use in context, and rated the pedagogical potential of those phrase-frames. Taken together, these studies have showcased strong associations between formulaic language and rhetorical moves as well as the different ways in which such associations may manifest in different genres and/or disciplines.

To date, the exploration of the connection between formulaic expressions and rhetorical moves in narrative stories remains an uncharted territory. Narratives, as a distinct genre, offer a rich and multifaceted structure that may interact with formulaic language in unique ways that are worthy of investigation. Additionally, there is a lack of research on lexical collocations in the body of genre analysis studies of the relationship between formulaic expressions and rhetorical moves in general. Given the importance of collocations in L2 learning and the challenges they pose to L2 learners discussed above, an analysis of the connection of this type of formulaic expression to rhetorical moves will not only contribute to an important research gap in genre analysis research on form-function connections but also yield findings that can help L2 learners better understand the contextual uses and communicative functions of collocations in narrative stories.

The present study

In light of the research gaps discussed above, this study aims to extract lexical collocations from a corpus of expert-written narrative stories, annotate the corpus with a new rhetorical move framework for narrative stories, and match the identified collocations to rhetorical moves in the corpus. Three research questions guide this study:

1. What are the frequently used lexical collocations in narrative stories?
2. What are the frequently employed rhetorical moves in narrative stories?
3. In what ways are lexical collocations connected to rhetorical moves in narrative stories?

Method

Corpus description

The target genre of the current study was narrative stories, which differ from other subgenres of stories with their 'problem-solution' plots (Martin and Rose 2008). We sourced narrative stories from *Reader's Digest*. Founded in 1922, *Reader's Digest* has cultivated a rich history of presenting narratives that resonate with readers across the globe; this historical legacy underscores its commitment to connecting with diverse audiences and maintaining a lasting impact on its readership (Reader's Digest 2022a). The magazine's thriving presence and global influence are also substantiated by recent data showing that the print and digital products of *Reader's Digest* now reach nearly one in every four adults in the USA, with millennials making up a third of its digital audience (CNBC 2019).

Along with its long history, wide reach, and ability to capture the interest of the digitally native generation, the abundance of expertly written narrative content on diverse topics in *Reader's Digest* makes it an appealing source of authentic materials for teaching and learning narrative stories. We collected 678 narrative stories published on the *Reader's Digest* website between September 2015 and February 2022. As shown in Table 1, the corpus contained 698,024 words, with an average of 1,029 words per story. These stories covered a conspicuous breadth of topics in 13 broad thematic domains (e.g. *True Stories*, *Pets & Animals*, *Knowledge*, *Humor*, etc.), encompassing a spectrum of human experiences such as love, loss, survival, and 'unbelievable coincidences' (Reader's Digest 2022b). This expansive thematic repertoire provides a rich terrain for linguistic investigation. For instance, the narratives falling within the thematic domain of *Knowledge* incorporate elements such as cognitive exercises, social etiquettes, facts, grammatical

Table 1: Description of the corpus

Topics	Number of texts (percentage)	Range	Mean number of words	Standard deviation	Mean number of paragraphs
True stories	454 (66.96%)	175–6171	1083.53	831.73	21.99
Pets & animals	48 (7.08%)	151–2,293	868.15	620.64	13.94
Knowledge	32 (4.72%)	168–2,559	981.25	667.22	18.13
Humor	6 (0.88%)	226–927	588.00	256.28	19.0
Money	5 (0.74%)	567–2,089	1560.40	545.13	29.4
Relationships	63 (9.29%)	188–3,572	1072.94	694.62	24.06
Home	1 (0.15%)	1,098–1,098	1098.00	0.0	23.0
Arts & entertainment	11 (1.62%)	422–2574	1152.73	611.98	24.36
Holidays & observances	40 (5.90%)	117–2,465	632.43	488.63	13.1
Travel	12 (1.77%)	247–2,399	933.00	644.07	17.83
Food	1 (0.15%)	772–772	772.00	0.0	11.0
Beauty & fashion	1 (0.15%)	380–380	380.00	0.0	6.0
Tech	4 (0.59%)	210–1,680	670.50	596.04	16.75
Total	678 (100%)	117–6,171	1029.53	779.62	20.83

and orthographic intricacies, as well as insights derived from psychology, while those categorized under the thematic domain *Humor* pivot toward lighthearted entertainment, characterized by the inclusion of cartoons, humorous photographs, comic stories, and jokes. Each story published by *Reader's Digest* was crafted by seasoned writers and underwent a rigorous review process before publication (Reader's Digest 2022a). As such, these stories not only afford language learners the opportunity to engage with authentic textual material that mirrors a wide array of human experiences but also exhibit a level of narrative writing expertise that can help learners develop their understanding of the linguistic and rhetorical features of narrative stories (Jiang 2022; Wang *et al.* 2023).

Extracting lexical collocations

To address the first research question, we extracted a list of lexical collocations from the corpus in five steps. Each collocation consisted of a node word lemma and a collocate word lemma, and only content word lemmas (i.e. nouns, verbs, adjectives, and adverbs) were considered as node words. First, we part-of-speech tagged and lemmatized each text in our corpus using the Stanford CoreNLP Toolkit (Manning *et al.* 2014). Second, we used the Keyword function of AntConc 4.0.6 (Anthony 2022) to compare the frequency of each lemma in our corpus with its frequency in the freely available lemma frequency list derived from the written section of the American National Corpus (ANC) (Reppen *et al.* 2005). The written section of the ANC contains 18,530,112 words and covers 11 written genres and is thus representative of general written English. The lemmas were ranked by their keyness value (based on the chi-squared measure) indicating how different their frequencies were in our corpus and the reference corpus. From the top 2,000 lemmas with the highest keyness values, we selected the content word lemmas (i.e. adjective, adverbs, nouns, and verbs) as node words (1,313 total). Third, we used the *Collocates* function of the online Corpus of

Contemporary American English (CoCA) (Davies 2020) to identify collocation of each node word. These were content word lemmas occurring within a five-word window on either side of the node word lemma at least 20 times per million words across at least 14 different texts, with a minimum mutual information value of 3 (Rogers *et al.* 2021). Fourth, the collocations extracted were filtered to remove repetitions. Specifically, if two identical collocations were retrieved using different node words, we removed the one retrieved with the lower-frequency node word. At the end of this step, a total of 622 unique lexical collocations that met the predetermined frequency and range thresholds were identified from the corpus. Given that the initial list was obtained using statistical criteria only, the first and fourth authors critically reviewed the extracted collocations to ensure the quality of the list and its pedagogical value. Specifically, following Rogers *et al.*' s (2021) criteria, the two researchers first independently judged whether each collocation was semantically meaningful and pedagogically relevant to the teaching and learning of English narrative stories. They then resolved all discrepancies through discussion. As a result of this process, collocations that contained proper nouns (e.g. *New/York*), or that were semantically vague (e.g. *think/do*), overly general (e.g. *people/by*), or overly domain specific (e.g. *digest/reader*) were excluded. The 219 lexical collocations judged by both researchers to be semantically meaningful and pedagogical useful were retained on the final list.

Identifying rhetorical moves

To answer the second research question, we developed a rhetorical move framework for narrative stories and annotated the stories in the corpus using this framework, following a three-phrase procedure summarized in Figure 1. In the first phase, the researchers collaboratively outlined three basic stages of narrative stories and a unique set of rhetorical moves for each stage based on an analysis of five stories randomly sampled from the corpus using Labov and Waletzky's (1967) framework of narratives, Martin and Rose's (2008) 'sections' of narratives, Flynn's (2018) functional stages for oral stories, as well as Jiang's (2022) and Wang *et al.*' s (2023) framework for classification of rhetorical moves and functions in narrative texts. A combination of top-down and bottom-up approaches was adopted in this phase, with 'an inductive analysis that forefronts cognitive judgement and produces a balance between the textual data and the researcher's knowledge of underlying rhetorical sections' (Cotos *et al.* 2017: 94). Each text was segmented into three stages based on content and general rhetorical purposes, and each stage was then segmented into moves of different functional types. The smallest unit of a stage was a paragraph, and the smallest unit of a rhetorical move was a sentence, in accordance with Holmes' (1997) definition of rhetorical moves.

Table 2 presents the three stages of narrative stories and their corresponding rhetorical moves included in our framework. The *Orientation* stage introduces the background of the story, expands the network of characters and causal relationships through minor events, and accounts for the situation. This stage includes three rhetorical moves, namely, *Setting the scene*, *Laying out the event*, and *Giving the current situation*. The *Incident* stage forms the main body of the narrative in terms of the plot by drawing forth the main conflict of the story and its gradual resolution. This stage

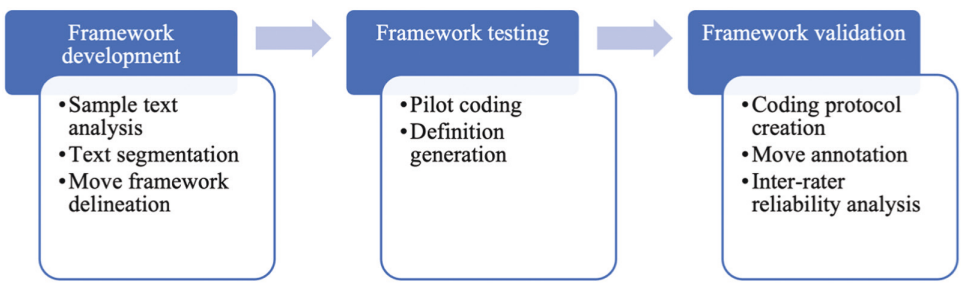

Figure 1: Move framework analytic procedure.

also includes three rhetorical moves, namely, *Complicating the story*, *Evaluating the status quo*, and *Dealing with the complication*. Finally, the *Conclusion* stage reveals the main idea of the story and gives an account of the ending of the story. This stage includes two rhetorical moves, namely, *Foregrounding the main idea* and *Ending the story*.

In the second phase, we invited nine coders to test the rhetorical move framework, including one college and one high school English teacher, three undergraduate English major students, two Master's students majoring in linguistics, and two doctoral students majoring in linguistics and educational technology, respectively. The first author presented the framework to the coders with definitions, examples, and indicative linguistic features for the moves. After familiarizing themselves with the framework, the coders each annotated the five stories selected in the first phase independently by highlighting text chunks with different colors corresponding to different rhetorical moves, as illustrated in Figure 2. The first author then met with all coders as a group to further clarify the definitions of the rhetorical moves, resolve discrepancies, discuss the distribution and linguistic features of the moves, and reflect on the applicability of the framework. The definitions of the rhetorical moves were fine-tuned as a result of this group discussion and reflection.

In the third phase, the first author and the nine coders annotated the remaining stories in the corpus using an online platform developed by a research assistant majoring in computer science (see Figure 3) and a corresponding coding protocol crafted by the researchers. In the platform, the coders focused on one paragraph (see the bolded paragraph in Figure 3) at a time. The sentences within the current paragraph were sequentially presented in the text box on the left one by one, and the annotator only needed to select the tag corresponding to the rhetorical move of the sentence and then click the 'submit' button to move on to the next sentence. Once the current paragraph was fully annotated, the next paragraph would automatically become available for annotation. The annotator could go back to any paragraph and revise the annotation of any sentence during the annotation process. At the end of the annotation process, the annotator would review the annotated text as a whole and make necessary changes.

Following a training session on the online interface, each coder was asked to annotate 67–68 texts. To assess inter-rater reliability, the 10 coders were divided into 5 pairs, and the 2 coders in

Table 2: Move framework for the annotation of narrative stories (based on Labov and Waletzky 1967; Martin and Rose 2008; Flynn 2018; adapted from Jiang 2022; Wang *et al.* 2023)

Stage	Move	Function
Orientation	Setting the scene	Introducing characters, time, location, and behavioral situation
Incident	Laying out the event	Presenting a temporal sequence of occurrences at the beginning of the text and giving an account of how one event leads to another
	Giving the current situation	Describing the temporary condition in the story before turning to the complication of the plot
	Complicating the story	Unfolding temporally ordered actions that lead to one or more crises
	Evaluating the status quo	Appraising complication by indicating attitudes, opinions, or the usuality of events, suspending or interwoven with action
	Dealing with the complication	Performing further actions to resolve the crisis
Conclusion	Foregrounding the main idea	Stating the main idea of the story after the resolution, acting as a culminating event
	Ending the story	Making a statement or comment about the story, or returning the narration to the present time

each pair annotated 10% (i.e. 7) of each other's texts. The five pairs achieved an average of 92.35% of agreement on the annotated moves, comparable to the 91.7% of agreement on move annotation reported by Lu *et al.* (2021b). The coders in each pair also checked each other's remaining annotated texts and resolved all discrepancies through discussion and consultation with the first author when necessary.

Connecting lexical collocations and rhetorical moves

To answer the third question, we matched each lexical collocation occurrence to the rhetorical move of the sentence containing it in the corpus. Additionally, we conducted correspondence analysis using SPSS 26.0 (IBM Corp. 2019) to explore the most prevalent lexical collocations in each rhetorical move. Correspondence analysis generates a graphical display of the rows and columns (in our case lexical collocations and rhetorical moves) of a contingency table that

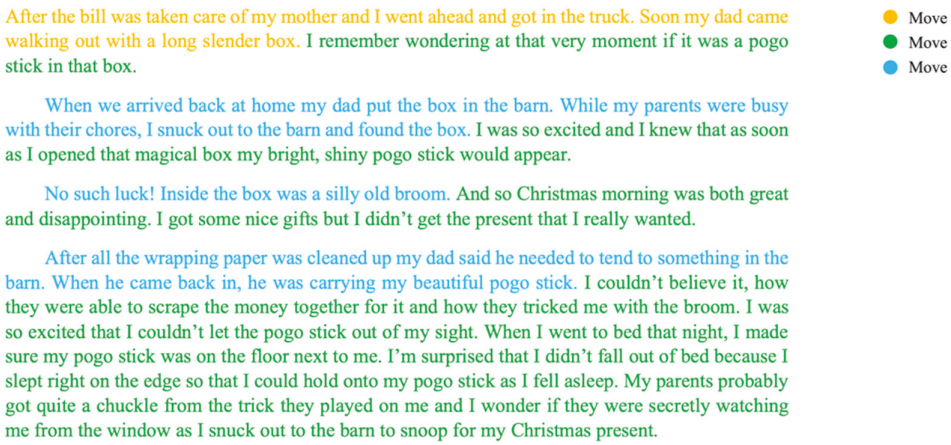

Figure 2: An excerpt of an annotated text.

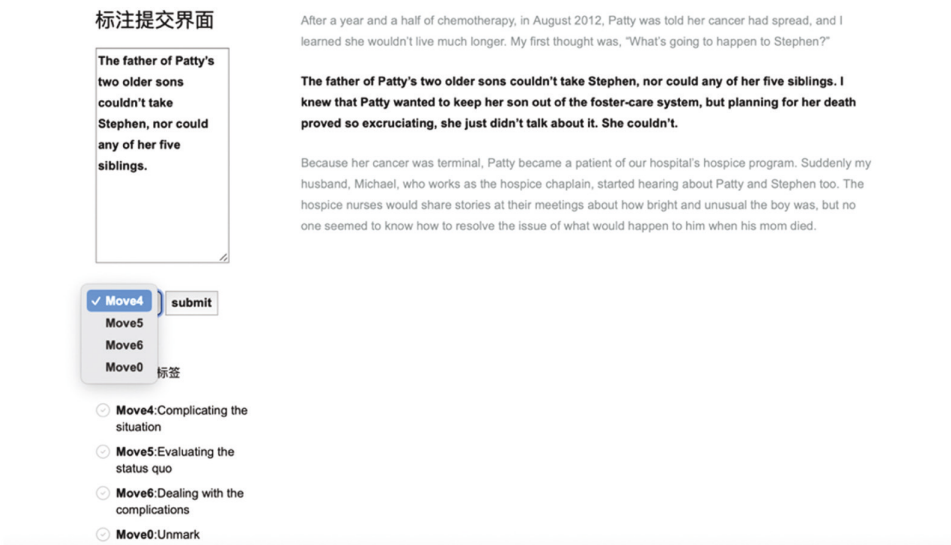

Figure 3: Online annotating platform.

visualizes the salient relationships among row-column pairs (in our case collocation-move pairs) (Mizumoto *et al.* 2017). The outcome of this analysis was examined to shed light on the associations between the collocations and rhetorical moves and the form-function connection in narrative stories in general.

Results

The lexical collocation list

Altogether, 219 lexical collocations were judged to be meaningful and pedagogically useful for the teaching and learning of English narrative stories. Table 3 presents the 10 most frequent lexical collocations along with example sentences extracted from the corpus. Supplementary Appendix A provides the complete lexical collocation list, with the frequency and range information and an example sentence for each collocation.

We examined the frequency of the word lemmas in the collocations using Cobb's (2018) *VocabProfiler*, which integrates the 25K most frequent English word families in the British National Corpus (BNC) and CoCA as reference lists (Nation 2012). As shown in Table 4, 186 of the 208 word lemma types (89.4%) in the lexical collocations extracted are from the first 1K most frequent word families in the BNC/CoCA, and the remaining 22 (10.6%) are from the second through fifth 1,000 most frequent word families. The high frequency of the word lemmas indicates that most of the collocations extracted should be relatively easily accessible to L2 English learners. However, as Rogers *et al.* (2021) pointed out, familiarity with high-frequency words in general English might not be 'sufficient for learners to comprehend or produce natural formulaic' English in specific genres due to genre variation (p. 153). As such, despite the high frequency of the word lemmas making up the lexical collocations extracted, it remains important to examine the uses and functions of these lexical collocations in the genre of narrative stories.

Table 3: Top 10 most frequent lexical collocations

Number	Node	Collocate	Example sentence
1	up	to	Helpless, he felt the water rising over his feet, then up to his knees, then his chest.
2	next	to	She would accept, sit next to him again, and that would be that, he thought.
3	time	at	Clara, 15 at the time, pressed the call button.
4	away	from	I stepped away from the wall.
5	home	in	Today, the couple are in their new home in Indianapolis.
6	get	out	'Let's get out of here', he told his wife, dressing quickly.
7	first	time	It's not the first time I've encountered them, nor will it be the last.
8	school	high	Jerry's shyness prevented him from saying a single word to her until their junior year in high school.
9	so	much	There is so much we cannot control, but we must try to make things better when we are able.
10	side	on	A car is driving slowly down the street on the other side.

While the collocations extracted are generally not structurally complete, their component word lemmas do exhibit strong grammatical correlations (Nation 2013). Inspired by Biber *et al.*' s (2004) categorization of lexical bundles, we categorized the collocations into six structural types, as summarized in Table 5. Type 1 collocations consist of verb phrase fragments (e.g. *tell/story*, *search/for*). Type 2 collocations incorporate noun phrase fragments (e.g. *young/woman*, *store/grocery*). Type 3 collocations are composed of adjective phrase fragments (e.g. *close/to*, *still/alive*). Type 4 collocations include dependent clause fragments (e.g. *now/by*, *when/finally*). Type 5 and Type 6 collocations include fragments of adverb phrases (e.g. *again/never, little/more*) and prepositional phrase fragments (e.g. *through/window*, *down/road*), respectively.

Rhetorical move framework of narrative stories

To address the second research question, we first developed a rhetorical move framework for narrative stories. This framework consisted of three stages, namely, *Orientation*, *Incident*, and *Conclusion*, which were further categorized into eight moves. We illustrate the rhetorical stages and moves in the framework with examples from different narrative stories from the corpus in this subsection and report on the distribution of these stages and moves in the corpus in the next subsection.

Table 4: Frequency distribution of the word lemmas in the lexical collocations extracted

Frequency level	Families (%)	Types (%)	Tokens (%)	Cumulative tokens (%)
K-1	186 (89.4)	186 (89.4)	415 (94.7)	94.7
K-2	14 (6.7)	14 (6.7)	15 (3.4)	98.1
K-3	4 (1.9)	4 (1.9)	4 (0.9)	99.0
K-4	3 (1.4)	3 (1.4)	3 (0.7)	99.7
K-5	1 (0.5)	1 (0.5)	1 (0.2)	100.0
Total	208 (100.0)	208 (100.0)	438 (100.0)	

Note. K-1, K-2, K-3, K-4, and K-5 denote the first, second, third, fourth, and fifth 1,000 most frequent word families, respectively.

Table 5: Structural types of lexical collocations

1. Verb phrase fragments (82 items)
For example, *tell/story*, *search/for*, *check/out*
2. Noun phrase fragments (41 items)
For example, *young/woman*, *seat/belt*, *store/grocery*
3. Adjective phrase fragments (7 items)
For example, *close/to*, *still/alive*
4. Dependent clause fragments (8 items)
For example, *now/by*, *when/finally*
5. Adverb phrase fragments (18 items)
For example, *again/once, again/never, little/more*
6. Prepositional phrase fragments (63 items)
For example, *through/window*, *down/road*, *day/end*

The first stage, *Orientation*, establishes the context and 'presents an expectant activity sequence' for the story (Martin and Rose 2008: 51). This stage consists of three moves. Move 1, *Setting the scene*, introduces the characters, time, location, and behavioral situation of the story. In Example (1), this move describes a character's routine behavioral situation before unfolding an unusual experience.

(1) *Almost every night for more than ten years, Kirk Alexander, 48, of Salem, Oregon ordered a late dinner from his local Domino's pizza store. He had no signature order—sometimes he would call for a salad, sometimes a pie, sometimes chicken wings.*
(Reader's Digest 2022c)

Move 2, *Laying out the event*, presents a temporal sequence of occurrences at the beginning of the text and gives an account of how one event leads to another. This move aims to expound on the context and acquaint the readers with the setting of the story by promoting a plot development. In Example (2), this move outlines an early exchange at the beginning of the story between a nurse and a patient, building up a friendly atmosphere while doing so.

(2) *Eventually I hit on a reliable way to make Patty smile: mentioning her nine-year-old son, Stephen. She'd tell me how well he did in school, how he'd been selected to read a poem he wrote at a local bookstore.*
(Reader's Digest 2022d)

Move 3, *Giving the current situation*, describes the temporary condition in the story before turning to the complication of the plot. This move helps readers conceive the overall environment and atmosphere, paving the way for the next stage. In Example (3), this move describes the environment the protagonist was in when he was hiking with his mother.

(3) *He watched 15 or so other hikers enjoy the vista; one hiker, around 60 and dressed in pink, was peeking over the lip of the precipice with her husband.*
(Reader's Digest 2022e)

The second stage, *Incident*, presents the main event of the story and signifies the major development of the plot. This stage consists of three moves. Move 4, *Complicating the story*, unfolds temporally ordered actions that break the normalcy established in the first stage and that lead to one or more conflicts of the story. In Example (4), this move describes a crisis from a pizza delivery driver's point of view, where a man who regularly ordered pizza did not come to the store and could not be found at home, disrupting the expected course of events.

(4) *He could plainly see that Alexander's TV set was on, as were his lights; but after several minutes, Alexander still didn't answer the door.*
(Reader's Digest 2022c)

Move 5, *Evaluating the status quo*, appraises the complication by indicating attitudes or opinions towards it. Evaluation is often 'deployed to suspend the action, increasing the narrative tension, and so intensifying the release when tension is resolved' (Martin and Rose 2008: 52). In Example (5), this move conveys the narrator's appraisal of and emotional reaction to the boy's situation, setting up the expectation for a resolution.

(5) *All of this, we knew, was tough medicine for such a little boy. We were total strangers to him, and he to us… 'Wow', I thought. 'This kid really is unusual'.*
(Reader's Digest 2022d)

Move 6, *Dealing with the complication*, responds to and resolves the crisis presented in Move 4. In Example (6), which immediately follows Example (5), this move describes the narrator's actions

in response to the boy's situation and a series of subsequent actions of the boy following the resolution of the crisis.

(6) *Finally, we took him home with us… All we could do was love him and pull him into the thick of our busy family life. He joined a basketball team.*
(Reader's Digest 2022d)

Stage three, *Conclusion*, consolidates the gist of the narrative and brings it to an end, returning the readers to the 'reality' of the present. This stage consists of two moves. Move 7, *Foregrounding the main idea*, states the main idea of the text after the resolution and reinterprets the central theme of the story either implicitly through a culminating event or explicitly through an advocation of the moral value of the story. In Example (7), this move unveils the central theme of the story through a final sequence of events.

(7) *On Christmas Eve, I learned that Stephen had never met Santa Claus, so I called the local mall to find an on-duty Santa Claus, shared a bit of our story, then rushed with my daughters to take Stephen to see him. Watching Santa hug Stephen close and have a beautiful heart-to-heart with him about his mother was one of the most moving moments of my life.*
(Reader's Digest 2022d)

Move 8, *Ending the story*, makes a statement or comment about the story or returns the narration to the present time. It may conclude the story by presenting the current situation after the main event, or reorient the story by presenting a new context that is not directly related to the main event. In Example (8), this move concludes the story by narrating a recent event, creating a new context that leaves the readers with much room for imagination.

(8) *Recently, the students in Stephen's class had to write a poem that started with 'Home is …' He was so sad because most of the other kids wrote poems that started with 'Home is my mom doing …' But he ended up writing this:*
Home is Karen listening to her inspirational meditation videos.
Home is Mike's delicious mac 'n' cheese.
Home is feeling cared for, loved, and protected.
(Reader's Digest 2022d)

Distribution of stages and moves

A total of 47,073 rhetorical chunks were annotated as one of the eight moves in the corpus. As shown in Table 6, moves of Stage 2, *Incident*, were the most predominant, occupying 71% of all moves. Within Stage 2, Move 4 was the most frequent, accounting for 29.14% of all moves (see also Figure 4), indicating the critical importance to detail the complication in the body of the story. Moves 6 and 5 were the second and third most frequent, accounting for 23.64% and 18.21% of all moves, respectively. The high proportion of these Stage 2 moves is in line with Martin and Rose's (2008) findings that the complication and resolution are the most essential moves in narrative stories. Moves of Stage 1, *Orientation*, occupied 17.83% of all moves. The most frequent move in this stage was Move 2, which accounted for 8.36% of all moves and 46.87% of all Stage 1 moves, indicating the importance of laying out the event at the beginning of the story. Finally, moves of Stage 3, *Conclusion*, accounted for the smallest proportion of all moves (i.e. 11.17%). Within this stage, Move 7 was more frequent than Move 8, indicating that storytellers often choose to highlight the main idea or theme of the story as they conclude the story. Move 8 was the least frequent among all moves, accounting for 4.16% of all moves only, consistent with Martin and Rose's (2008) finding that the ending of a story tends to be brief or optional.

Connection between lexical collocations and rhetorical moves in narrative stories

The results of the correspondence analysis of collocation-move association are summarized in Figure 5 using a two-dimensional scatterplot. The full collocation–move matrix and the dimensional score of each collocation are provided in Supplementary Appendices B and C, respectively. The scatterplot allows us to visually examine the connection between lexical collocations and moves. Specifically, collocations that are more frequently used to realize a particular move can

Table 6: Distribution of moves in the corpus

Move	Number	Percent
Stage 1, *Orientation*	8,391	17.83
Move 1: Setting the Scene	2,321	4.93
Move 2: Laying Out the Event	3,933	8.36
Move 3: Giving the Current Situation	2,137	4.54
Stage 2, *Incident*	33,421	71.00
Move 4: Complicating the Story	13,718	29.14
Move 5: Evaluating the Status Quo	8,573	18.21
Move 6: Dealing with the Complication	11,130	23.64
Stage 3, *Conclusion*	5,261	11.17
Move 7: Foregrounding the Main Idea	3,302	7.01
Move 8: Ending the Story	1,959	4.16
Total	47,073	100.0

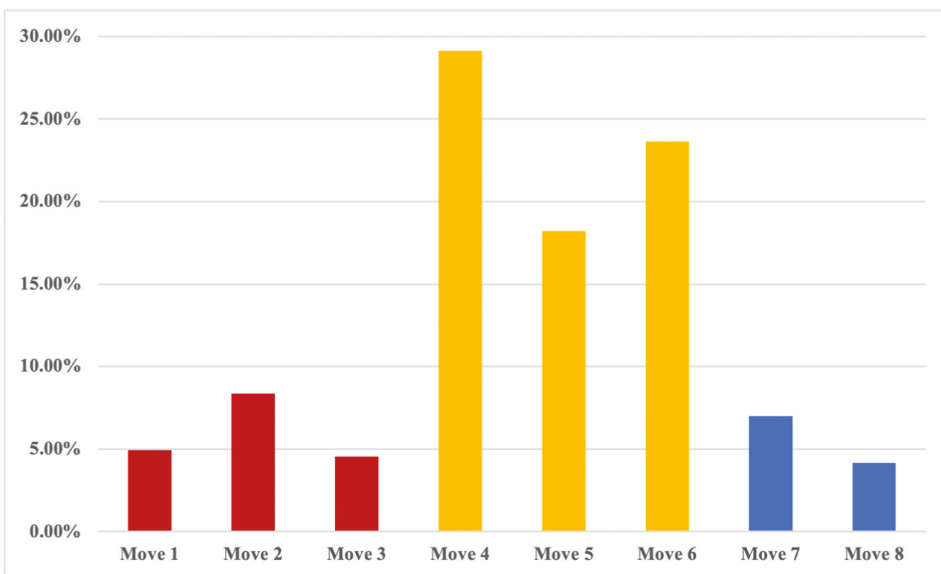

Figure 4: Percentages of different moves in the corpus.

be found closer to that move than to others. Many of the frequent lexical collocations are closely clustered around the origin, where Moves 2, 4, 5, 6, and 8 are located, indicating that a large variety of the collocations are involved in laying out the event, developing the plot, and ending the story in the narratives. However, Moves 1 and 7 and to a lesser extent Move 3 are farther away from the origin and are each associated with a smaller range of collocations. For instance, Move 1, which is placed on the top of the figure, is close to such collocations as *away/home*, *long/before*, *school/high*, and *police/department*, indicating that these collocations are strongly associated with scene setting in the narrative stories in the corpus.

Following the correspondence analysis aimed at establishing the connection between collocations and rhetorical moves, we conducted a qualitative analysis of the functions of the collocations, guided by the premise that this connection is inherently contingent upon the functions these collocations serve within the narrative discourse (Cortes 2013). This analysis involved a systematic examination of the collocations in their specific contexts of use to determine the broader communicative purposes they help fulfill in the narratives. This detailed analysis yielded three overarching functions that encapsulate the varied roles these collocations play in advancing narrative stories.

The first function that lexical collocation was frequently used to realize was the introduction of situational information, such as time (e.g. *day/before*, *few/ago*), location (e.g. *through/window*, *across/country*), and behavioral situation (e.g. *up/grow*, *first/time*), as illustrated in Examples (9) to

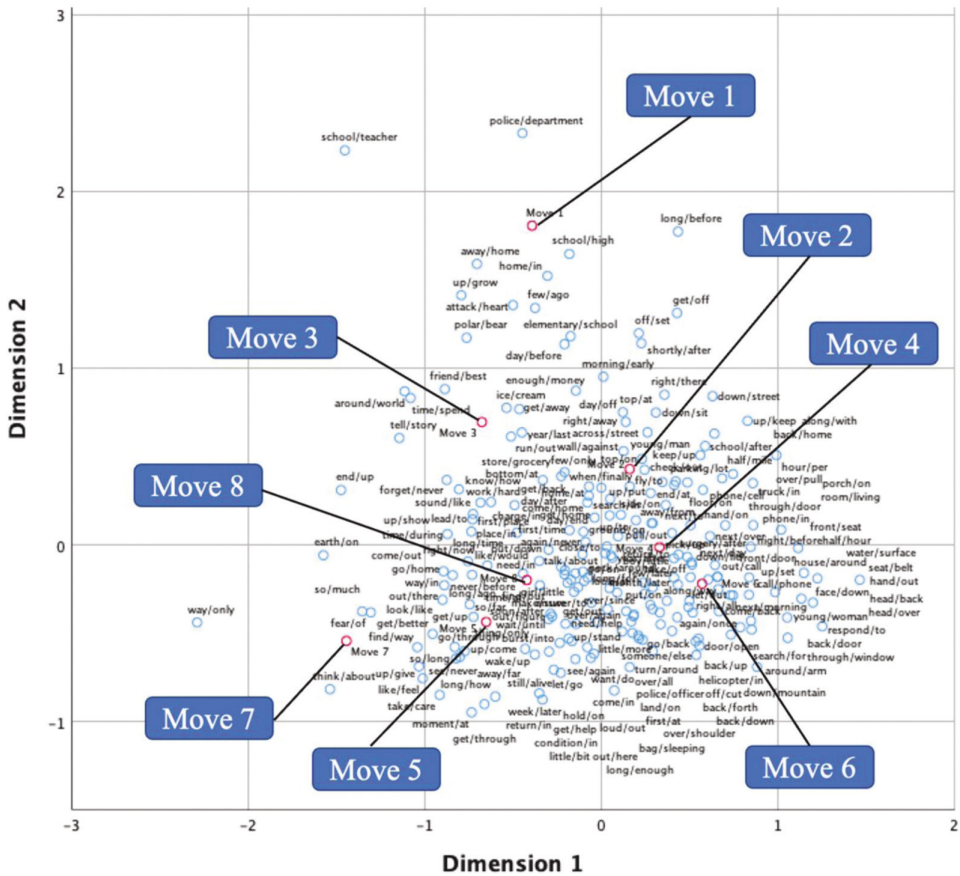

Figure 5: Scatterplot of collocation-move correspondence.

(11). This function was found to be particularly pronounced in the three Stage 1 moves, where background information is of paramount importance.

(9) *By now it is past noon and even hotter than the* ***day before****.* (Time introduction)
(10)*He then peered* ***through*** *the front* ***window****.* (Location introduction)
(11)*The kittens are all* ***grown up*** *now.* (Behavioral situation introduction)

The second function pertains to the substantial role that collocations play in advancing the storyline, predominantly evident in Stage 2 moves. This is not surprising, given that plot development is the most important part of a narrative story (Martin and Rose 2008). The three specific functions collocations were used to realize include the elaboration of events (e.g. *soon/after*, *neck/around*), the description of circumstances (e.g. *like/feel*, *want/do*), and resolution actions (e.g. *hold/on*, *get/help*). Event elaboration collocations often involve nouns or adverbs to highlight aspects of the plot development, as illustrated in Example (12), circumstance description collocations often involve mental state verbs, as illustrated in Example (13), and collocations describing resolution actions tend to involve the use of action verbs, as illustrated in Example (14). The use of diverse lexical collocations in these ways allowed the narrators to effectively convey event sequences and character actions as the narratives unfolded.

(12) *In the process, the backpack strap wrapped* ***around*** *his* ***neck****, strangling him.* (Event elaboration)
(13)*And sometimes they* ***feel*** *a bit* ***like*** *a miracle.* (Circumstance description)
(14)*Hello, I'm trying to* ***get*** *some* ***help*** *to you.* (Resolution action)

The third function centers on the expression of attitudes from the characters' perspectives (e.g. make/sure, think/about) and the statement of the main idea of the story from the narrator's stance (e.g. end/up, so/much), as illustrated in Examples (15) and (16). While the event evaluation and circumstance description functions shared some of the similar collocations, the former emphasized the subjective feelings of the characters about the event, while the latter emphasized the surrounding environment itself.

(15) *All of this has made me* ***think about*** *what a wonderful life I have lived.* (Event evaluation)
(16)*And sometimes you* ***end up*** *being the answer to their prayers.* (Main idea statement)

Discussion

The analysis for our first research question centered on identifying a list of lexical collocations that can be useful for the teaching and learning of English narrative stories. Following the recent practice in corpus-based formulaic language research (e.g. Cortes 2013; Omidian *et al.* 2018), we extracted a candidate list of 622 collocations from a corpus of narrative stories using multiple statistical criteria. Informed by Rogers *et al.*' s (2021) observation that relying on corpus data alone is 'insufficient in creating usable teaching and learning materials' (154), we manually filtered the list to ensure that all collocations included in our final list were meaningful and pedagogically useful, consistent with other studies that aimed at creating pedagogically useful resources from text corpora (e.g. Durrant 2009; Lu *et al.* 2018; Rogers *et al.* 2021). The majority of the component words of the collocations were within the first 1,000 most frequent words in the BNC/CoCA, indicating that most of the collocations extracted from the narrative stories are in the general domain and should be accessible or teachable to beginning or intermediate L2 English learners. Given the importance of context in understanding and learning collocations, we have provided example sentences from the corpus for the collocations extracted (see Supplementary Appendix A). Informed by Biber *et al*'.s (2004) structural taxonomy for lexical bundles, we categorized the lexical collocations on our list into four types of structural fragments, including

verb phrase fragments, noun phrase fragments, adjective phrase fragments, and adverb phrase fragments. It became clear later that collocations of different structural types tended to be more frequently used for different functions.

The analysis for our second research question involved devising a rhetorical move framework for narrative stories, annotating the corpus of narrative stories using the framework, and examining the distribution of different rhetorical stages and moves in the corpus. This analysis helps us understand the rhetorical structure or strategical organization (Cotos *et al.* 2017) of narrative stories produced by expert writers. Martin and Rose (2008) proposed three stages for narrative stories, that is, *Orientation*, *Complication*, and *Resolution*. Building on their proposal and based on our analysis of the corpus of narrative stories, we proposed the following three stages of narrative stories: *Orientation*, *Incident,* and *Conclusion*, and, more importantly, fleshed out a set of moves with distinct rhetorical functions for each stage. An analysis of the distribution of the stages and moves in the corpus showed a primary rhetorical emphasis on Stage 2, with Move 4 (*Complicating the story*) ranked as the most frequent among all moves, in line with Martin and Rose's (2008) findings that complication and resolution make up the most essential part of a narrative story. In Stages 1 and 3, Move 2 (*Laying out the event*) and Move 7 (*Foregrounding the main idea*) were the most frequent, respectively, shedding light on what narrators pay the most attention to in the *Orientation* and *Conclusion* stages of the story. The focus on foregrounding the main idea in the *Conclusion* stage resonates with Deane *et al.* 's (2019) finding pertaining to the emphasis on the morals and themes of the story in the *Narrative Reflection* phase of storytelling.

The analysis for our third research question included a correspondence analysis of lexical collocations and rhetorical moves in narrative stories, followed by an analysis of the specific communicative functions of the lexical collocations used in different rhetorical moves in context. No collocation occurred exclusively in any one move, different from findings that some larger phraseological units such as lexical bundles and phrase-frames are exclusively move-specific in academic writing (e.g. Cortes 2013; Lu *et al.* 2021b). However, the correspondence analysis showed that Moves 1, 3, and 7 have the strongest association with a relatively small range of lexical collocations, while the other five moves, including all three Stage 2 moves, have been realized using a large variety of lexical collocations. This difference may have arisen from the differences in the degree of specificity and/or uniqueness of the rhetorical functions of the moves. For example, Move 1 entails substantially greater use of collocations that introduce the time, location, and behavioral situation of the story to set the scene than other moves, while the three Stage 2 moves involve a broader range of collocations to cover different aspects of the plot to develop the story. A further analysis of the collocations in context revealed three types of communicative functions of the collocations aligned with the rhetorical functions of the three stages of narrative stories, namely, to introduce situational information, advance the storyline, and express attitudes. Collocations with these different functions also tended to be of different structural types. The patterns of collocation-move associations unveiled by the correspondence analysis lend support for the theoretical framework of genre competence that highlights the nexus between language features and rhetorical strategies in specific genes in general (e.g. Swales 1990; Tardy 2016; Lu *et al.* 2021a) and the notion that such linguistic choices in narratives serve as vehicles for specific discourse functions in particular (Trosborg 1997).

Our findings hold good promise for enhancing genre-based pedagogy. The move framework proposed for narrative stories can be adopted and adapted by teachers to facilitate L2 learners' understanding of the general rhetorical structure of the narrative story genre and the specific ways in which expert storywriters piece together a well-structured narrative story (Jiang 2022). Teachers can also draw on our findings regarding the distribution of rhetorical stages and moves to help L2 learners note the relative frequency of chunks with different rhetorical functions and appreciate how expert storywriters allocate spaces for different rhetorical moves. The list of lexical collocations extracted from the corpus and the example sentences that accompany the list could be used as resources to help the learners expand their repertoire of commonly used collocations in the narrative story genre. The structural categorizations of the

collocations on the list could make it easier for teachers and learners to focus on collocations of specific structural types if desired. More importantly, our findings regarding the connection between lexical collocations and rhetorical moves and the primary functions that lexical collocations are frequently used to realize can help learners gain a more nuanced understanding of how expert storytellers employ different linguistic devices to materialize different communicative functions.

To facilitate the pedagogical application of our findings, we have set up a searchable online interface[1] that allows teachers and learners to search the corpus for collocations by the node word, retrieve a list of collocations along with their frequency in the corpus, the rhetorical stages and moves they occur in, and example sentences from the corpus they occur in (Wang *et al.* 2023). The online interface can serve as an especially useful tool for learners to explore lexical collocations in relation to rhetorical moves in narrative stories. This exploration could complement teacher-designed genre analysis activities to promote L2 English learners' competence in deploying appropriate collocations for different rhetorical moves in narrative story writing (e.g. Dong and Lu 2020), in line with the growing recognition that corpus-informed language instruction can help EFL learners internalize language patterns in meaningful communicative contexts (Rodríguez-Fuentes and Swatek 2022).

Conclusion

Extending the emerging line of genre analysis research on bridging the form-function gap into the understudied genre of narrative stories, this study has identified a list of frequent lexical collocations from a corpus of narrative stories, developed a rhetorical move framework for narrative stories, examined the distribution of rhetorical stages and moves in the corpus, and explored the connection between lexical collections and rhetorical moves in narrative stories. Our findings have useful implications for the theoretical framework of genre competence and genre-based pedagogy. Importantly, the findings of the study culminated in a pedagogically useful online interface for searching the corpus for collocations by the node word and exploring their use in sentences realizing different rhetorical stages and moves in context.

Several important avenues for future research exist. First, the specific pedagogical value of our research findings and the online search interface for genre-based narrative writing pedagogy can be explored in classroom teaching contexts. Second, a parallel analysis of the frequently used collocations, the distribution of rhetorical stages and moves, and the connection between collocations and rhetorical moves in a corpus of narrative stories produced by L2 English learners would shed light on the gaps in their collocational and rhetorical knowledge of this genre. Finally, the scope of analysis of the form-function connection in this genre can be expanded to other types of phraseological units and other linguistic constructions in general.

Supplementary data

Supplementary material is available at *Applied Linguistics* online.

Notes on contributors

Yishi Jiang is a PhD student at Shanghai Jiao Tong University. His research interests include language assessment, corpus linguistics and computer-assisted language learning.

Xiaofei Lu is professor of Applied Linguistics and Asian Studies at The Pennsylvania State University. His research interests are in corpus linguistics, English for Academic Purposes, second language acquisition, and second language writing. He is the author of *Corpus Linguistics and Second Language Acquisition: Perspective, Issues, and Findings* (2023, Routledge).

Fengkai Liu is a PhD student in the Department of Linguistics and Translation at City University of Hong Kong. His research interest primarily focuses on computational linguistics.

Jianxin Zhang is an MA student at University of International Business and Economics. Her research interests include corpus linguistics and language testing.

Tan Jin (corresponding author) is a professor in the School of Foreign Studies at South China Normal University. His research interests include corpus linguistics, language assessment and computer-assisted language learning. His research has appeared in *Language Learning & Technology*, *Language Testing*, *Modern Language Journal* and *TESOL Quarterly Address for correspondence*: School of Foreign Studies, South China Normal University, Guangzhou, Guangdong, China. <tjin@scnu.edu.cn>

Note

[1] The URL of the online searchable interface is https://www.languagedata.net/writer/search/.

References

Ackermann, K. and Y. Chen. 2013. 'Developing the academic collocation list (ACL): A corpus-driven and expert-judged approach,' *Journal of English for Academic Purposes* 12/4: 235–47. https://doi.org/10.1016/j.jeap.2013.08.002

Anthony, L. 2022. *AntConc. Computer Software*. Waseda University.

Berman, R. A. and B. Nir-Sagiv. 2007. 'Comparing narrative and expository text construction across adolescence: A developmental paradox,' *Discourse Processes* 43: 79–120. https://doi.org/10.1207/s15326950dp4302_1

Biber, D., S. Conrad, and V. Cortes. 2004. 'If you look at…: Lexical bundles in university teaching and textbooks,' *Applied Linguistics* 25: 371–405. https://doi.org/10.1093/applin/25.3.371

Bychkovska, T. and J. J. Lee. 2017. 'At the same time: Lexical bundles in L1 and L2 university student argumentative writing,' *Journal of English for Academic Purposes* 30: 38–52. https://doi.org/10.1016/j.jeap.2017.10.008

Casal, J. E. and M. Kessler. 2020. 'Form and rhetorical function of phrase-frames in promotional writing: A corpus-and genre-based analysis,' *System* 95: 102370. https://doi.org/10.1016/j.system.2020.102370

CNBC. 2019. 'How Reader's Digest went from life support to modern, digital-first multimedia brand', available at https://www.cnbc.com/2020/03/02/how-readers-digest-transformed-into-a-digital-multimedia-brand.html. Accessed 27 August 2023.

Cobb, T. 2018. VocabProfiler VP Compleat BNC-COCA-25. *Computer Software*. Available at https://www.lextutor.ca/vp/comp/. Accessed 1 September 2022.

Conrad, S. and D. Biber. 2004. 'The frequency and use of lexical bundles in conversation and academic prose,' *Lexicographica: International Annual for Lexicography* 20: 56–71.

Cortes, V. 2013. 'The purpose of this study is to: Connecting lexical bundles and moves in research article introductions,' *Journal of English for Academic Purposes* 12/1: 33–43. https://doi.org/10.1016/j.jeap.2012.11.002

Cotos, E., S. Huffman, and S. Link. 2017. 'A move/step model for methods sections: Demonstrating rigour and credibility,' *English for Specific Purposes* 46: 90–106. https://doi.org/10.1016/j.esp.2017.01.001

Davies, M. 2020. The Corpus of Contemporary American English (CoCA). *Data Set*. Available at https://www.english-corpora.org/coca/. Accessed 1 September 2022.

Deane, P., *et al*. 2019. 'The key practice, building and sharing stories and social understandings: The intrinsic value of narrative,' *Educational Testing Service* 2019/1: 1–78. https://doi.org/10.1002/ets2.12266

Dong, J. and X. Lu. 2020. 'Promoting discipline-specific genre competence with corpus-based genre analysis activities,' *English for Specific Purposes* 58: 138–54. https://doi.org/10.1016/j.esp.2020.01.005

Durrant, P. 2009. 'Investigating the viability of a collocation list for students of English for academic purposes,' *English for Specific Purposes* 28/3: 157–69. https://doi.org/10.1016/j.esp.2009.02.002

Ellis, N. C. 2012. 'Memory for language' in P. Robinson (ed.): *Cognition and Second Language Instruction*. Cambridge University Press, pp. 32–68.

Flynn, E. E. 2018. 'Storying experience: Young children's early use of story genres,' *Text & Talk* 38/4: 457–80. https://doi.org/10.1515/text-2018-0010

Holmes, R. 1997. 'Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines,' *English for Specific Purposes* 16/4: 321–37. https://doi.org/10.1016/S0889-4906(96)00038-5

Hyland, K. 2008. 'Genre and academic writing in the disciplines,' *Language Teaching* 41/4: 543–62.

IBM Corp. 2019. *IBM SPSS Statistics for macOS. Computer Software*. IBM Corp.

Jiang, Y. 2022. '*The connection between collocations and rhetorical moves in narrative stories: A corpus-based study*', unpublished Master's thesis, Sun Yat-sen University.

Jović, S. 2020. 'The narrative games we play: Varied use of narrative strategies across genres and socioeconomic positions,' *Language & Communication* 74: 87–102. https://doi.org/10.1016/j.langcom.2020.06.001

Kang, J. Y. 2005. 'Written narratives as an index of L2 competence in Korean EFL learners,' *Journal of Second Language Writing* 14/4: 259–79. https://doi.org/10.1016/j.jslw.2005.10.002

Kanoksilapatham, B. 2007. 'Rhetorical moves in biochemistry research articles' in D. Biber, U. Connor, and T. A. Upton (eds): *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. John Benjamins Publishing Company, pp. 73–119.

Labov, W. and J. Waletzky. 1967. 'Narrative analysis: Oral versions of personal experience' in J. Helm (ed.): *Essays on the Verbal and Visual Arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*. University of Washington Press, pp. 12–44.

Lawler, S. 2002. 'Narrative in social research' in T. May (ed.): *Qualitative Research in Action*. SAGE Publications, pp. 242–58.

Lu, X., J. E. Casal, and Y. Liu. 2021a. 'Towards the synergy of genre- and corpus-based approaches to academic writing research and pedagogy,' *International Journal of Computer Assisted Language Learning and Teaching* 11/1: 59–71.

Lu, X., J. Yoon, and O. Kisselev. 2021b. 'Matching phrase-frames to rhetorical moves in social science research article introductions,' *English for Specific Purposes* 61: 63–83. https://doi.org/10.1016/j.esp.2020.10.001

Lu, X., J. Yoon, and O. Kisselev. 2018. 'A phrase-frame list for social science research article introductions,' *Journal of English for Academic Purposes* 36: 76–85. https://doi.org/10.1016/j.jeap.2018.09.004

Manning, C. D., *et al*. 2014. 'The Stanford CoreNLP Natural Language Processing Toolkit'. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 55–60.

Martin, J. R. and D. Rose. 2008. *Genre Relations: Mapping Culture*. Equinox.

Mizumoto, A., S. Hamatani, and Y. Imao. 2017. 'Applying the bundle–move connection approach to the development of an online writing support tool for research articles,' *Language Learning* 67/4: 885–921. https://doi.org/10.1111/lang.12250

Nation, P. 2013. *Learning Vocabulary in Another Language*. Cambridge University Press.

Negretti, R. and L. McGrath. 2018. 'Scaffolding genre knowledge and metacognition: Insights from an L2 doctoral research writing course,' *Journal of Second Language Writing* 40: 12–31. https://doi.org/10.1016/j.jslw.2017.12.002

Omidian, T., H. Shahriari, and A. Siyanova-Chanturia. 2018. 'A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts,' *Journal of English for Academic Purposes* 36: 1–14. https://doi.org/10.1016/j.jeap.2018.08.002

Reader's Digest. 2022a. 'About Reader's Digest', available at https://www.rd.com/about-readers-digest/. Accessed 19 January 2022.

Reader's Digest. 2022b. 'Real life stories', available at https://www.rd.com/real-life-stories/. Accessed 19 January 2022.

Reader's Digest. 2022c. 'That time ordering Domino's pizza saved a man's life', available at https://www.rd.com/article/pizza-saved-mans-life/. Accessed 31 January 2022.

Reader's Digest. 2022d. 'A nurse's story: taking in my cancer patient's 9-year-old son was the best thing I ever did', available at https://www.rd.com/article/nurse-cancer-patient-son/. Accessed 2 February 2022.

Reader's Digest. 2022e. 'How an 18-year-old helped save a hiker who fell 75 feet off a mountain cliff', available at https://www.rd.com/article/teenager-saves-hiker/. Accessed 3 February 2022.

Reppen, R., N. Ide, and K. Suderman. 2005. American National Corpus (ANC) Second Release. *Data Set*. Available at https://anc.org/. Accessed 1 September 2022.

Rodríguez-Fuentes, R. A. and A. M. Swatek. 2022. 'Exploring the effect of corpus-informed and conventional homework materials on fostering EFL students' grammatical construction learning,' *System* 104: 102676. https://doi.org/10.1016/j.system.2021.102676

Rogers, J., *et al*.. 2021. 'The creation and application of a large-scale corpus-based academic multi-word unit list,' *English for Specific Purposes* 62: 142–57. https://doi.org/10.1016/j.esp.2021.01.001

Sanford, A. J. and C. Emmott. 2012. *Mind, Brain and Narrative*. Cambridge University Press.

Shin, D. and P. Nation. 2008. 'Beyond single words: The most frequent collocations in spoken English,' *ELT Journal* 62/4: 339–48. https://doi.org/10.1093/elt/ccm091

Swales, J. 1981. *Aspects of Article Introductions*. Language Studies Unit.

Swales, J. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.

Tardy, C. M. 2016. *Beyond Convention: Genre Innovation in Academic Writing*. University of Michigan Press. https://doi.org/10.3998/mpub.5173647

Trosborg, A. 1997. 'Text typology: Register, genre and text type' in A. Trosborg (ed.): *Text Typology and Translation*. Benjamins Translation Library, pp. 3–24.

Underwood, G., N. Schmitt, and A. Galpin. 2004. 'The eyes have it: An eye-movement study into the processing of formulaic sequences' in N. Schmitt (ed.): *Formulaic Sequences*. John Benjamins, pp. 153–72.

Wang, Y., *et al*. 2023. 'Connecting form with function: Model texts for bilingual learners' narrative writing,' *Assessing Writing* 57: 100753. https://doi.org/10.1016/j.asw.2023.100753

Webb, S. and E. Kagimoto. 2011. 'Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning?,' *Applied Linguistics* 32/3: 259–76. https://doi.org/10.1093/applin/amq051

Wray, A. and M. Perkins. 2000. 'The functions of formulaic language: An integrated model,' *Language & Communication* 20/1: 1–28. https://doi.org/10.1016/S0271-5309(99)00015-4

Yoon, H. J. 2021. 'Challenging the connection between task perceptions and language use in L2 writing: Genre, cognitive task complexity, and linguistic complexity,' *Journal of Second Language Writing* 54: 100857. https://doi.org/10.1016/j.jslw.2021.100857