

**A Corpus-Based Analysis of Lexical Bundles in Traditional and Modern English
Fairy Tales****Balqista Surya Azizah¹ ✉, Aprillia Firmonasari²**^{1,2}Fakultas Ilmu Budaya, Universitas Gadjah Mada, Indonesia

Article Info	Abstract
<i>Article History:</i> Received 30 March 2026 Approved 24 April 2026 Published 30 April 2026	Lexical bundles are sequences of words that frequently recur and play a key role in shaping natural language use. Despite extensive research on lexical bundles in academic and spoken registers, their use in literary texts, particularly fairy tales, remains underexplored. As a genre rooted in oral storytelling, fairy tales are characterized by linguistic patterns that may reflect the use of such bundles across different periods. This study aims to identify four-word lexical bundles in traditional and modern English fairy tales, focusing on their frequency, structure, and function. Using a corpus-based descriptive-comparative approach, two balanced corpora of 400 traditional and 400 modern fairy tales were analyzed, comprising approximately 2.53 and 2.55 million words, respectively. The bundles were extracted using AntConc based on frequency and dispersion criteria. In both corpora, noun phrase fragments, prepositional phrase fragments, and referential expressions were the most dominant patterns, indicating that lexical bundles mainly function to construct narrative settings and organize story events. These results suggest that traditional fairy tales rely more heavily on formulaic language, while modern fairy tales exhibit greater linguistic variation. Overall, the findings indicate both continuity and subtle change in the use of lexical bundles across the two periods.
Keywords corpus linguistics, fairy tales, formulaic language, lexical bundles	

© Copyright 2026

INTRODUCTION

Formulaic sequence is one of the characteristics of language use in everyday interactions. Instead of relying solely on single words or isolated phrases, communication frequently consists of recurrent patterns of words that create naturalness in expression (Chen et al., 2023; Pawley & Syder, 1983; Wood, 2015). Wray (2002) explains that the use of a multi-word sequence is greatly influenced by two main factors, such as limitations in language processing and communicative pressures, which make formulaic sequences easier to be stored in long-term memory and shaped by what is considered normal or appropriate in their community.

As cited by Gray and Biber (2015), sequences of words used together repeatedly in predictable ways have been extensively studied over the past three decades, as a significant portion of everyday language consists of such recurring sequences rather than strictly compositional units. These sequences have been studied under various labels, one of which is lexical bundles.

The concept of lexical bundles was initially proposed by Biber et al. (1999), where they examined their occurrence across conversational and academic writing. They define lexical bundles as sequences of words that frequently co-occur within a particular register. Moreover, lexical bundles acts as recurring expressions, regardless of whether they are idiomatic or not, and without

✉ Corresponding author:

E-mail: balqistasuryaazizah@mail.ugm.ac.id

p-ISSN: 2252-6323

e-ISSN: 2721-4540

considering their structural status. Examples of such bundles include *as a result of*, *on the other hand*, and *I don't know if*.

Numerous studies on lexical bundles have mainly concentrated on professional and academic contexts, such as academic writing, legal texts, and spoken discourse. In the realm of academic writing, Hyland (2008) examined four-word bundles within a corpus comprising research articles, master's theses, and doctoral dissertations across various disciplines, and conveyed that the discipline influenced variations in lexical bundles. Meanwhile, in legal texts discourse, Goźdz-Roszkowski (2020) examined the use of lexical bundles in judicial opinions and reported that, in legal texts, these bundles are primarily employed to promote legal certainty. In spoken discourse, Alkhelaiwi (2023) studied lexical bundles in English podcasts from Saudi Arabia and found that the dominant type of bundles was interpersonal. Despite this extensive attention in academic contexts, studies on lexical bundles in literary genres remain limited, particularly in the context of fairy tales. Therefore, this study aims to address this gap by analyzing lexical bundles in literary genres, particularly fairy tales.

The concept of a fairy tale is hard to pin down. Zipes (2001) points out that while there are countless fairy tales, defining them as a single genre is challenging because they have been described in so many different ways. A strictly literal interpretation might limit fairy tales to those that involve fairies (Briggs, 1976). However, this narrow definition falls short for understanding fairy tales in the 20th and 21st centuries, as the term has expanded to include a broad range of stories over time (Haase, 2008).

As a form of folk literature initially spread through oral media, fairy tales have played a substantial role in conveying moral values and cultural representations from pre-modern times to the present. Since the 17th century, fairy tales have been documented in written form by folklore collectors, from Charles Perrault in France, Jacob and Wilhelm Grimm in Germany to Joseph Jacobs and Andrew Lang in English-speaking countries (Bottigheimer, R. B., 2002; Classen, A., 2021; Teverson, A., 2016). Meanwhile, in the 20th and 21st centuries, modern fairy tales have emerged through the works of authors such as Roald Dahl,

Neil Gaiman, and Kate DiCamillo, who have reconstructed classic fairy tales with a more complex and critical approach (Greenhill & Magnus-Johnston, 2023).

To ensure conceptual clarity, this study adopts explicit definitions of traditional and modern fairy tales grounded in previous studies. The terms traditional and modern fairy tales in this study refer to the definitions provided by Fernández (2019), Lester (2015), Sala-Suszyńska (2021), and Shahrin (2023), where traditional fairy tales are those compiled before the 19th century, while modern fairy tales are those created from the 20th century onward. The main difference between traditional and modern fairy tales lies in their historical context and the themes they explore. Traditional fairy tales, which rooted in oral folklore, tend to focus on moral lessons, clear distinctions between good and evil, and supernatural elements. On the other hand, modern fairy tales reflect changes in society which feature more complex characters and ambiguous morals. They often address contemporary issues such as gender roles, social justice, and individual identity. This shift shows a move from straightforward narratives to more layered and nuanced storytelling.

As stated by Lazim (2015), one of the characteristics of linguistic richness in fairy tales is the use of recurring language patterns, including lexical bundles. Several examples of lexical bundles in fairy tales are provided by Kopecká (2016), who identified common opening phrases in English fairy tales, such as *once upon a time*, *there lived a*, and *in a faraway land*.

In this context, the present study focuses on lexical bundles in traditional and modern English fairy tales by examining their frequency, structure, and function. By comparing these two types of texts, the study seeks to identify the use of lexical bundles across the two types of texts. To examine this phenomenon, a strong data-based approach is needed to study the lexical bundles objectively and systematically.

To address this issue, corpus linguistics offers a suitable framework. Corpus linguistics is an approach for studying language that uses real language data sets (corpus) as the basis for analysis (McEnery & Hardie, 2012). This approach allows researchers to identify linguistic patterns based on

the frequency, distribution, and context of their use in texts. By employing computational tools such as AntConc (Anthony, 2024), researchers are able to find N-grams, which are the most commonly occurring sequences of three words (3-grams), four words (4-grams) or more.

This study offers a new approach to understanding fairy tales. It broadens the scope of study on lexical bundles, which has largely been dominated by academic and formal genres (Martínez, 2022). Given the differences in linguistic characteristics between traditional and modern fairy tales that have been explained previously, the study of lexical bundles in these two types of texts is important for understanding language dynamics in popular literary genres. In addition, this study also contributes a new perspective to the field of corpus linguistics, especially in the context of formulaic language analysis, which is often overlooked in literary genres.

METHODS

A corpus-based method was adopted to address the aims of the study. This study employs a descriptive-comparative design. The descriptive approach describes the characteristics of lexical bundles as they occur within the text. Meanwhile, the comparative approach compares the frequency, structural patterns, and functional uses of bundles in the traditional and modern English fairy tales. The data for this study were lexical bundles in traditional and modern English fairy tales that were collected from publicly accessible online repositories and digital archives of English fairy tales.

To ensure methodological transparency and replicability, the texts were selected based on explicit criteria proposed by Sinclair (2005). The selection criteria focused on language, source, and publication period. Only original English fairy tales from English-speaking countries were included in the corpus, while texts translated from other languages were excluded. The corpus consisted of written fairy tale texts in the form of books and short stories, including electronic versions obtained from legally accessible online sources such as public digital archives and authorized websites that provide open access to

literary materials. Finally, publication period was used to distinguish the two groups, with traditional fairy tales documented before the 19th century and modern fairy tales published from the 20th century onward.

The bundles examined consist of only four words (4-grams), in line with Biber and Barbieri (2007), who contend that research on English lexical bundles typically centers on sequences of four words. This is because three-word bundles are too numerous and considered less interesting, whereas five- and six-word bundles are comparatively rare. Moreover, longer bundles often incorporate shorter sequences within them.

The collected data were arranged into two corpora, namely the traditional English fairy tales corpus (TEFT) and the modern English fairy tales corpus (MEFT). Each corpus was compiled into .txt format to be compatible with the analysis software. To ensure data validity, the corpora were cleaned to include only the main narrative content, with non-linguistic elements that may interfere with the analysis, such as page numbers, footnotes, author/editor names, and illustrations, removed. Each corpus was designed to have an equivalent size to ensure comparability and validity between the two types of texts. The TEFT corpus consists of 400 individual texts and contains 2,582,465 tokens and 32,682 types. Meanwhile, the MEFT corpus comprises 400 individual texts with 2,586,692 tokens and 40,386 types.

Each corpus was imported into AntConc version 4.3.1 (Anthony, 2024), which was used to conduct frequency-based analyses. Following Hyland (2008), a frequency threshold of 20 occurrences per million words was applied. Moreover, the dispersion criterion was generally established as being present in minimum five texts in the corpus (Biber et al., 2004).

Using the N-gram tool in AntConc, the lists of lexical bundle were then extracted from both the TEFT and MEFT corpora to obtain their frequency and distribution. The extraction results were exported into Microsoft Excel, organized into separate sheets for each corpus, and sorted by frequency of occurrence from highest to lowest. The Collocates tool was employed to identify the words that appear before and after the bundle, and further supported by the Keyword in Context

(KWIC) tool to examine the contextual usage of the bundles.

The log-likelihood test was employed to determine whether the differences in the frequencies, structural patterns, and functional uses of lexical bundles between the TEFT and MEFT corpora were statistically significant. In this analysis, the log-likelihood values were calculated for both lexical bundle types and tokens to identify differences in the frequency of occurrence across the two corpora.

Trustworthiness was ensured through triangulation by combining multiple data sources, namely traditional and modern fairy tales, analytical tools (N-gram, KWIC, and Collocate in AntConc), and peer review to improve the accuracy and objectivity of the findings.

Data analysis was conducted in three main interconnected stages, namely frequency, structure, and function analysis. These stages were applied to both corpora to identify and compare the use of lexical bundles present in each corpus.

RESULTS AND DISCUSSION

This section presents the findings of the study, namely how the frequency, structure, and function of lexical bundles vary between traditional and modern English fairy tales. It begins with an overview of the frequency of lexical bundles in the TEFT and MEFT. It then examines the characteristics of the bundles to analyze their variation. Finally, it compares the bundles in traditional and modern corpora to identify the similarities and differences across the two periods.

Overall Frequency of Lexical Bundles in Traditional and Modern English Fairy Tales

The analysis identified 108 lexical bundle types (8,646 tokens) in the TEFT corpus and 66 types (5,873 tokens) in the MEFT corpus which

demonstrates that lexical bundles are widely used in both traditional and modern fairy tales. Table 1 displays the number of tokens, types, and the standardized type-token ratio (STTR) of the two corpora. Tokens represents the total word count in the corpus, whereas types refer to the number of distinct words. Lexical diversity was assessed using the standardized type-token ratio (STTR), which normalizes variation by calculating the type-token ratio per 1,000 words.

Table 1. Summary of Tokens, Types, and Standardized Type-Token Ratio (STTR) in TEFT and MEFT

	Token	Type	STTR (%)
Traditional	2,582,465	32,682	39
Modern	2,586,692	40,386	44

Based on the STTR values, the two corpora show comparable levels of lexical diversity. The difference in STTR values (39% vs. 44%) indicates a slightly higher lexical variation in the MEFT corpus. However, the relatively small difference reflects a tendency for modern texts to be more lexically diverse. The following section examines the characteristics of the most frequently used lexical bundles in each corpus to further clarify this pattern of variation.

High-Frequency Lexical Bundles in Traditional and Modern English Fairy Tales

This section outlines the highest-frequency lexical bundles in the TEFT and MEFT. The bundles are ranked by frequency, and the range reflects the number of texts in which each bundle occurs. The ten highest-frequency lexical bundles in the TEFT and MEFT are displayed in the following table.

Table 2. Ten Highest-Frequency Lexical Bundles in TEFT and MEFT

TEFT				MEFT			
No	Type	Freq	Range	No	Type	Freq	Range
1	at the end of	193	80	1	the edge of the	252	79
2	the top of the	186	81	2	the rest of the	189	70
3	in the middle of	184	90	3	in the middle of	178	65
4	the middle of the	174	85	4	the end of the	175	60
5	the rest of the	158	44	5	on the other side	166	48
6	the king of the	150	35	6	for the first time	159	66
7	the end of the	148	63	7	the other side of	159	52

p-ISSN: 2252-6323
e-ISSN: 2721-4540

8	in the midst of	146	51	8	the side of the	156	35
9	of the king of	131	37	9	the top of the	155	54
10	on the other side	130	60	10	the middle of the	150	53

As shown in Table 2, the bundle *at the end of* ranks first in the TEFT with 193 occurrences across 80 texts, while the tenth-ranked bundle, *on the other side*, occurs 130 times in 60 texts. Meanwhile, the bundle *the edge of the* ranks first in the MEFT, with 252 occurrences across 79 texts. The tenth-ranked bundle, *the middle of the*, appears 166 times across 48 texts.

Although the frequencies of the bundles in each corpus are close, there is a notable difference between the highest-ranked and the tenth-ranked bundles. This variation shows that the distribution of bundle frequencies is not evenly balanced. Moreover, the TEFT and MEFT corpora also share several identical bundles. This overlap suggests that certain lexical bundles are consistently used across both traditional and modern fairy tales. While both sets of texts rely on some similar formulaic patterns, each corpus shows its own pattern of recurring expressions.

Shared Bundles in Traditional and Modern English Fairy Tales

The analysis of the Traditional English Fairy Tales (TEFT) and Modern English Fairy Tales (MEFT) corpora reveals that 31 identical lexical bundles occur in both corpora. These bundles are referred to as shared lexical bundles. These shared bundles, again, show that both traditional and modern fairy tales often rely on similar recurring language patterns, regardless of their period.

The shared bundles in both corpora consist of a range of lexical constructions from phrases to clauses. Among the 31 shared bundles, 23 are structurally incomplete, while eight are complete. In other words, most bundles act as fragments of larger syntactic units that need extra complements to form complete sentences. This finding is in line with Hyland and Jiang (2019) and Pan et al. (2016) who state that lexical bundles are part of bigger structural unit.

When categorized by their grammatical form, phrasal bundles dominate the shared bundles with 23 occurrences, while clausal bundles are less common with only eight occurrences (*and there was*

a, as if he had, but there was no, he was going to, i don't know what, i don't want to, what do you mean, what do you think). Among the 23 phrasal bundles identified in the shared bundles, several types of phrases are found, including 11 noun phrases (*other side of the, the back of the, the bottom of the, the door to the, the edge of the, the end of the, the middle of the, the other side of, the rest of the, the side of the, the top of the*) and 12 prepositional phrases (*at the back of, at the bottom of, at the end of, at the same time, once upon a time, at the top of, for a long time, for the first time, in front of the, in the middle of, on the other side, out of the way*).

An analysis of the frequency of shared lexical bundles across both corpora reveals differences in which bundles occur most often. As an example, the most frequent shared bundle in the TEFT, is *at the end of*, compared with the highest in the MEFT, *the edge of the*, serve a similar function in expressing limits or boundaries. The bundle *at the end of* (TEFT=193; MEFT=131) is generally used to indicate the final point of a location or an event. In contrast, the bundle *the edge of the* (TEFT=92; MEFT=252) is used to describe the outer boundary or border of a place or object. This shows a difference in the preferred spatial expressions across the two corpora. This pattern suggests that spatial expressions play a central role in structuring narrative descriptions across both periods.

Structure of Lexical Bundles in Traditional and Modern English Fairy Tales

The lexical bundles identified in the TEFT and MEFT corpora exhibit two types of structure, namely complete and incomplete. Complete bundles are those that are not fragmented and can form a well-structured phrase, such as *once upon a time, the king and queen, took a deep breath*, or a clause, such as *he did not know, she shook her head, what do you think*. In contrast, incomplete bundles are structurally fragmented, resulting in phrases or clauses that are not fully realized due to the absence of one or more essential elements. For example, the bundle *the center of the* is considered incomplete because it does not constitute a complete phrase unless it is followed by an element

such as the center of the room. Similarly, the bundle *are you going to* is classified as an incomplete clause, as it is fragmented and terminates at the word *to* and leaves the structure unfinished.

According to the core elements that compose it, in line with the lexical bundles structure of Biber et al. (2004), the bundles in the TEFT and MEFT corpora can be categorized into:

1) Verb phrase fragments, 2) Dependent clause fragments, 3) Noun phrase fragments, 4) Prepositional phrase fragments, 5) Comparative expressions, 6) Other expressions. To provide a clearer comparison of the overall structure of lexical bundles in both corpora, Table 3 presents the proportion of the bundles in the TEFT and MEFT datasets.

Table 3. Structural Distribution of Lexical Bundles in TEFT and MEFT

Structure	Types				Tokens			
	TEFT	%	MEFT	%	TEFT	%	MEFT	%
(connector +) 1 st /2 nd person pronoun + VP fragment	8	7.41	3	4.55	567	6.56	215 ⁻⁻⁻⁻	3.66
(connector +) 3rd person pronoun + VP fragment	18	16.67	7 ⁻	10.61	1248	14.43	454 ⁻⁻⁻⁻	7.73
Discourse marker + VP fragment	0	0	0	0	0	0	0	0
Verb phrase (with non-passive verb)	3	2.78	2	3.03	173	2	134 ⁻	2.28
Verb phrase with passive verb	0	0	0	0	0	0	0	0
yes-no question fragments:	0	0	2	3.03	0	0	111 ⁺⁺⁺⁺	1.89
WH-question fragments	3	2.78	4	6.06	192	2.22	325 ⁺⁺⁺⁺	5.53
Verb phrase fragments	32	29.63	18⁻	27.27	2180	25.21	1239⁻⁻⁻⁻	21.1
1 st /2 nd person pronoun + dependent clause fragment	1	0.93	2	3.03	66	0.76	137 ⁺⁺⁺⁺	2.33
WH-clause fragments	2	1.85	0	0	121	1.4	0 ⁻⁻⁻⁻	0
If-clause fragments	1	0.93	0	0	54	0.62	0 ⁻⁻⁻⁻	0
(verb/adjective+) to-clause fragment	3	2.78	0 ⁻	0	188	2.17	0 ⁻⁻⁻⁻	0
That-clause fragments	6	5.56	0 ⁻⁻	0	400	4.63	0 ⁻⁻⁻⁻	0
Dependent clause fragments	13	12.04	2⁻⁻	3.03	829	9.59	137⁻⁻⁻⁻	2.33
(connector +) Noun phrase with <i>of</i> -phrase fragment	22	20.37	15	22.73	2257	26.1	1921 ⁻⁻⁻⁻	32.71
Noun phrase with other post-modifier fragment	0	0	1	1.52	0	0	51 ⁺⁺⁺⁺	0.87
Other noun phrase expressions	3	2.78	1	1.52	199	2.3	60 ⁻⁻⁻⁻	1.02
Noun phrase fragments	25	23.15	17	25.76	2456	28.41	2032⁻⁻⁻⁻	34.6
Prepositional phrase fragments	29	26.85	25	37.88	2518	29.12	2179⁻⁻⁻⁻	37.1
Comparative expressions	1	0.93	0	0	53	0.61	0⁻⁻⁻⁻	0
Other expressions	8	7.41	4	6.06	610	7.06	286⁻⁻⁻⁻	4.87
Total	108	100	66⁻⁻	100	8646	100	5873⁻⁻⁻⁻	100

LEGEND:

(----) Statistically significant underuse in MEFT corpus (at $p < 0.0001$, critical value 15.13)

(--) Statistically significant underuse in MEFT corpus (at $p < 0.001$, critical value 10.83)

(-) Statistically significant underuse in MEFT corpus (at $p < 0.01$, critical value 6.63)

(-) Statistically significant underuse in MEFT corpus (at $p < 0.05$, critical value 3.84)

(+++++) Statistically significant overuse in MEFT corpus (at $p < 0.0001$, critical value 15.13)

(+++ Statistically significant overuse in MEFT corpus (at $p < 0.001$, critical value 10.83)

(++ Statistically significant overuse in MEFT corpus at $p < 0.01$, critical value 6.63)

(+ Statistically significant overuse in MEFT corpus (at $p < 0.05$, critical value 3.84)

Based on the log-likelihood calculation and its interpretation, the distribution presented in Table 3 indicates that all six structural categories are represented in both corpora. All these major structural categories occur more frequently in the TEFT corpus than in the MEFT corpus, both in terms of bundle types and token frequencies. The frequency and distribution of each category vary with verb phrase fragments, noun phrase fragments, and prepositional phrase fragments constitute the most prominent categories. Meanwhile, dependent clause fragments, comparative expressions, and other expressions occur less frequently. This distribution suggests that lexical bundle use is selective rather than uniform.

Among all categories, verb phrase fragments are the most prominent in the TEFT corpus, with statistically significant higher frequency than in MEFT based on the log-likelihood value. This dominance suggests that traditional fairy tales strongly rely on action-oriented constructions, especially bundles built around third-person pronouns such as *he went to the*, *he said to the*, and *and there was a*. The higher frequency of third-person pronoun-based bundles in the TEFT corpus can be explained by the narrative style of traditional fairy tales. These narratives are typically told from an external perspective, allowing the narrator to describe characters and events from outside the story world (Biber et al., 2004; see also Jahn, 2026). Cited from Kaczorowska et al., (2025), such a perspective enables a broader coverage of events across the narrative. This makes third-person constructions particularly suitable for storytelling. In contrast, the more limited range of bundles in the MEFT reflects a shift toward less formulaic and more diverse narrative styles that subvert traditional structures.

A similar contrast appears in dependent clause fragments, which are significantly more frequent in TEFT than in MEFT in both types and tokens. The higher frequency of dependent clause fragments in TEFT indicates that traditional fairy tales make greater use of embedded structures to

connect actions and events within the narrative. Bundles such as *that he could not*, *if it had been*, and *to take care of* help maintain narrative continuity by linking clauses in predictable ways. In modern fairy tales, the sharp reduction of these bundles suggests a stylistic shift toward more compact and less formulaically dependent narration.

Noun phrase fragments and prepositional phrase fragments are highly frequent in both corpora. Noun phrase fragments are largely represented by bundles such as *the top of the*, *the edge of the*, and *the king of the*, while prepositional phrase fragments include bundles such as *on the other side*, *at the same time*, and *in the middle of*. These structures contribute to the descriptive texture of fairy tales by identifying characters and locating objects within the story world.

The categories of comparative expressions and other expressions occur much less frequently in both corpora. This indicates that they are not central structural resources in fairy tale discourse. Comparative expressions are represented only by the bundle *as well as the* in TEFT. Likewise, other expressions include heterogeneous forms such as *every now and then*, *as soon as the*, and *as if it were*, which function mainly to express hypothetical situations. Although this structure occurs in both corpora, the bundles are significantly more frequent in TEFT, again reflecting stronger dependence on recurring formulaic expressions.

Overall, the structural findings show that traditional fairy tales exhibit greater structural repetition across all major bundle categories which highlights a stronger dependence on formulaic narrative constructions. Modern fairy tales retain the same structural patterns but use them less frequently, suggesting a movement toward more varied linguistic expression.

Function of Lexical Bundles in Traditional and Modern English Fairy Tales

The lexical bundles identified in this study were classified according to the functional taxonomy proposed by Biber et al. (2004). This framework was chosen instead of the taxonomy developed by Hyland (2008) because the functional

categories proposed by Biber et al. were found to be more suitable for categorizing the bundles identified in the present corpora. It should be noted that the classification of lexical bundles into functional categories and subcategories was not intended to provide a definitive count, but rather to serve as an interpretive framework for understanding how the bundles function in context. Since the function of a lexical bundle may vary depending on its usage, each bundle was

examined in its original context before being assigned to a functional category.

Drawing on the functional taxonomy proposed by Biber et al. (2004), the lexical bundles identified in the TEFT and MEFT corpora were classified into four major functional categories: 1) Stance expressions, 2) Discourse organizers, 3) Referential expressions, 4) Special functions. The distribution of these four functional categories in the TEFT and MEFT corpora is presented in Table 4.

Table 4. Functional Distribution of Lexical Bundles in TEFT and MEFT

Structure	Types				Tokens			
	TEFT	%	MEFT	%	TEFT	%	MEFT	%
Epistemic	7	6.48	2	3.03	445	5.15	137 ⁻⁻⁻⁻	2.33
Attitudinal/modality	12	11.11	9	13.64	859	9.94	563 ⁻⁻⁻⁻	9.59
Stance Expressions	19	17.59	11	16.67	1304	15.08	700⁻⁻⁻⁻	11.92
Topic Introduction/Focus	3	2.78	2	3.03	265	3.07	158 ⁻⁻⁻⁻	2.69
Topic Elaboration / Clarification	21	19.44	9 ⁻	13.64	1488	17.21	676 ⁻⁻⁻⁻	11.51
Discourse Organizers	24	22.22	11⁻	16.67	1753	20.28	834⁻⁻⁻⁻	14.2
Identification/Focus	13	12.04	4 ⁻	6.06	1000	11.57	231 ⁻⁻⁻⁻	3.93
Imprecision	0	0	0	0	0	0	0	0
Specification of Attributes	6	5.56	0 ⁺⁻	0	421	4.87	0 ⁻⁻⁻⁻	0
Time/Place/Text Reference	42	38.88	39	59.08	3934	45.50	3997	68.06
Referential Expressions	61	56.48	43	65.14	5355	61.94	4228⁻⁻⁻⁻	71.99
Politeness	0	0	0	0	0	0	0	0
Simple Inquiry	1	0.93	1	1.52	54	0.62	111 ⁺⁺⁺⁺	1.89
Reporting	3	2.78	0 ⁻	0	180	2.08	0 ⁻⁻⁻⁻	0
Special Functions	4	3.71	1	1.52	234	2.7	111⁻⁻⁻⁻	1.89
Total	108	100	66⁺⁻	100	8646	100	5873⁻⁻⁻⁻	100

LEGEND:

(----) Statistically significant underuse in MEFT corpus (at $p < 0.0001$, critical value 15.13)

(---) Statistically significant underuse in MEFT corpus (at $p < 0.001$, critical value 10.83)

(--) Statistically significant underuse in MEFT corpus (at $p < 0.01$, critical value 6.63)

(-) Statistically significant underuse in MEFT corpus (at $p < 0.05$, critical value 3.84)

(+++++) Statistically significant overuse in MEFT corpus (at $p < 0.0001$, critical value 15.13)

(+++)

(++) Statistically significant overuse in MEFT corpus at $p < 0.01$, critical value 6.63)

(+) Statistically significant overuse in MEFT corpus (at $p < 0.05$, critical value 3.84)

Table 4.1 shows the distribution of lexical bundle functional categories in the TEFT and MEFT corpora. All functional categories appear in both corpora. Among the four major categories, referential expressions are the most dominant in

both corpora. Referential expressions are used to identify entities and to provide references to time, place, or textual elements. The findings indicate that more than half of the lexical bundles identified in both corpora are used primarily to refer to

entities, locations, and time references. The high frequency of referential expressions in both corpora is closely related to the nature of fairy tales as narrative texts, which rely heavily on lexical bundles to identify the characters and describe the settings. The log-likelihood result shows that referential expressions are statistically overused in TEFT in MEFT. This suggests that traditional fairy tales rely more on repetitive referential patterns in the narrative.

Discourse organizers constitute the second most frequent functional category in both corpora and are significantly more frequent in TEFT than in MEFT. Bundles in this function aim to structure the flow of discourse by introducing, elaborating, or clarifying information within the text. Their higher frequency in traditional fairy tales suggests a stronger reliance on recurring formulaic sequences to maintain narrative flow. From a functional perspective, this pattern reflects the role of lexical bundles as discourse-building devices in orally rooted narratives, where recurrent expressions help structure the flow of events in ways that are predictable and easy to follow. In contrast, the lower frequency in modern fairy tales indicates that similar organizational functions are achieved through more varied linguistic constructions rather than repeated formulaic sequences.

Similarly, stance expressions occur more frequently in TEFT than in MEFT. These bundles convey the speaker's attitudes, feelings, intentions, or judgments toward a proposition. The relatively lower frequency of this category in both corpora suggests that the primary role of lexical bundles in fairy tales is not to express evaluation or personal stance, but to support the progression of the narrative. Although stance expressions are relatively infrequent, their presence in TEFT indicates that lexical bundles occasionally support the characters' intentions or attitudes which is often through repeated formulaic patterns. This difference is also statistically significant, with stance expressions being overused in TEFT. This finding is consistent with Biber et al. (2004), who found that stance expressions are more characteristic of interactive academic registers such as classroom teaching, where speakers frequently convey personal attitudes, certainty, and obligation. Since fairy tales primarily focus on

narrating events rather than expressing interpersonal stance, this functional category appears less frequently than referential expressions and discourse organizers.

The last subfunction, special functions are the least frequent category in both corpora. This subfunction includes bundles that serve interactional purposes such as politeness, simple inquiry, and reporting. Their low frequency in both corpora indicates that such interpersonal functions are not central to the use of lexical bundles in fairy tale narratives. Instead, lexical bundles are used mainly for narrative and descriptive purposes. Although this category is also statistically overused in TEFT, its overall proportion remains very limited compared to the other functional categories. This is largely due to the greater frequency of special function bundles in conversational registers, where speakers rely on formulaic expressions to manage real-time interaction.

Overall, the functional distribution of lexical bundles in both corpora shows that fairy tales mainly use lexical bundles to support narrative description rather than interpersonal interaction. The dominance of referential expressions, followed by discourse organizers, suggests that these bundles play an important role in identifying characters, describing settings, and organizing the sequence of events in the story. This finding is in line with Biber et al. (2004), who found that referential bundles are common in written registers because they help structure information clearly. However, while referential bundles in academic writing are mainly used to present information precisely, in fairy tales they are more closely related to building the narrative world through references to place, time, and characters.

These distributional patterns indicate that lexical bundles in fairy tales serve primarily referential functions, whereas traditional fairy tales exhibit a stronger dependence on recurrent formulaic sequences than modern fairy tales. These findings suggest that although the functional categories identified by Biber et al. (2004) appear in fairy tales, the way these bundles function is shaped by the narrative purpose of the genre. As previous studies have also mentioned, different registers employ a distinctive set of bundles

according to their use across different contexts (Biber et al., 2004; Li & Schmitt (2023).

CONCLUSION

This study examined the frequency, structural patterns, and functional distributions of lexical bundles in traditional and modern English fairy tales. The findings reveal that lexical bundles are widely used in both corpora, but occur more frequently in traditional fairy tales. This indicates that traditional narratives rely more heavily on recurrent formulaic sequences, whereas modern fairy tales employ more varied linguistic patterns. In both corpora, the most prominent structural categories were verb phrase fragments, noun phrase fragments, and prepositional phrase fragments, while referential expressions and discourse organizers dominated the functional distribution. These results indicate that lexical bundles in fairy tales serve primarily as narrative resources that help construct settings, organize events, and maintain textual coherence.

The comparison between the two corpora reveals an important stylistic difference in the use of formulaic language. Traditional fairy tales rely more heavily on recurrent structural and functional patterns, reflecting the formulaic nature of oral storytelling traditions in which repetition supports continuity and predictability. By contrast, modern fairy tales show a reduced dependence on recurring bundles and greater lexical variation, suggesting a movement toward more flexible narrative expression.

Theoretically, this study broadens the application of Douglas Biber et al.'s framework by demonstrating that lexical bundles in fairy tales fulfill genre-specific narrative functions rather than the argumentative or stance-related functions often observed in academic discourse. These findings contribute to corpus-based research on formulaic language by showing that the use of lexical bundles reflects broader developments in narrative style, particularly the transition from oral-based formulaic narration to more varied literary expression.

REFERENCES

- Alkhelaiwi, S. (2023). Lexical bundles in a Saudi general-audience podcast in English: A corpus analysis. *World Journal of English Language*, 13(2), 355–363.
<https://doi.org/10.5430/wiel.v13n2p355>
- Anthony, L. (2024). *AntConc (Version 4.3.1) [Computer Software]*. Tokyo, Japan: Waseda University.
<https://www.laurenceanthony.net/software/AntConc>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405.
<https://doi.org/10.1093/applin/25.3.371>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Bottigheimer, R. B. (2002). Misperceived Perceptions: Perrault's Fairy Tales and English Children's Literature. *Children's Literature*, 30, 1–18.
- Briggs, K. (1976). *An encyclopedia of fairies*. New York, NY: Pantheon Books.
- Chen, K., Gu, L., & Bai, Q. (2023). Processing Chinese formulaic sequences in sentence context: a comparative study of native and non-native speakers. *Humanities and Social Sciences Communications*, 94(1), 1–9.
<https://doi.org/10.1057/s41599-023-02137-4>
- Classen, A. (2021). The Fairy Tales by the Brothers Grimm and their Medieval Background. *The German Quarterly*, 94(2), 165–175.
<https://doi.org/10.1111/gequ.12173>
- Fernández, A.S. (2019). *The Metamorphosis of Fairy Tales during the 20th and 21st Centuries*. [Bachelor's Thesis, Valladolid University]. Valladolid University Repository.
- Goźdz-Roszkowski, S. (2020). Communicating dissent in judicial opinions: A comparative, genre-based analysis. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, 33, 381–401. <https://doi.org/10.1007/s11196-020-09711-v>
- Gray, B. & Biber, D. (2015). Phraseology. In D. Biber & R. Reppen, *The Cambridge Handbook of English Corpus Linguistics*.
- Greenhill, P., & Magnus-Johnston, A. (2023). *The Fairy Tale in Contemporary Contexts: Retelling, Reimagining, and Rewriting*. University Press of Mississippi.
- Haase, D. (2008). *The Greenwood encyclopedia of folktales and fairy tales*. Westport, CT: Greenwood Press.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*,

- 27(1), 4–21.
<https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K., & Jiang, F. (2019). Points of Reference: Changing Patterns of Academic Citation. *Applied Linguistics*, 40, 64–85.
<https://doi.org/10.1093/applin/amx012>
- Jahn, M. (2026). *Narratology 3.01: A Guide to the Theory of Narrative*. English Department, University of Cologne. Accessed from www.uni-koeln.de/~ame02/pppn.pdf
- Kaczorowska, K., Conduit, J., & Goodman, S. (2025). Evoking altruistic behaviours: The effects of storytelling in different narrative modes. *Australasian Marketing Journal*, 1–24.
<https://doi.org/10.1177/14413582251389231>
- Kopecká, M. (2016). *The Use of “Once Upon a Time” in a Corpus of Fairy Tales and in The British National Corpus*. [Bachelor’s Diploma Thesis, Masaryk University]. Masaryk University Repository.
- Lazim, A. (2015). A Fairy-Tale Collection in Central London. *iBBYlink*, (43), 24–26.
- Lester, A. J. (2015). *The Impact of Traditional and Modern Fairy Tales on Society and Its Individuals*. [Bachelor’s Thesis, University of Arizona]. University of Arizona Repository.
- Martínez, R. A. (2022). Teaching lexical bundles in context: Insights from EFL learners’ discourse awareness. *Language Teaching Research*, 26(4), 657–675.
<https://doi.org/10.1177/1362168820980659>
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics*. Cambridge University Press.
<https://doi.org/10.1017/cbo9780511981395>
- Pan, F., Reppen, R. & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60–71.
<https://doi.org/10.1016/j.jeap.2015.11.003>
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). Longman.
- Sala-Suszyńska, J. (2021). Fairy Tales – Yesterday and Today. *Prima Educatione*, 5, 193–204.
<http://dx.doi.org/10.17951/pe.2021.5.193-204>
- Shahrin, T. (2023). *Reincarnation of the Myths: The Evolution of Fairy Tales in Modern Literature*. [Master’s Thesis, BRAC University]. BRAC University Repository.
- Sinclair, J. (2005). Corpus and Text — Basic Principles. In M. Wynne (Eds.), *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books.
- Teverson, A. (2016). The Fairy-Tale Collections of Andrew Lang and Joseph Jacobs: Identity, Nation, Empire. *Gramarye*, 9, 7–17.
- Wood, D. (2015). *Fundamentals of Formulaic Language: An Introduction*. London: Bloomsbury.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Zipes, J. (2001). *Sticks and stones: The troublesome success of children’s literature from Slovenly Peter to Harry Potter*. Routledge.