Identification of Emotion Words in Popular Children's Literature and Their Application to Semantic-based DDL

Kunihiko Miura
The University of Shimane, Faculty of International Relations, International Communication Department
k-miura@u-shimane.ac.jp

Abstract

What kind of emotion words do children's literature authors use in their works and which are children exposed to through encounters with such literature? This study identifies such emotion words and develops their pedagogical use as semantic-based data driven learning (DDL) for young learners of English as a Foreign Language (EFL). RQ1 What word levels do the authors use in popular children's literature? RQ2 What high frequency emotion words are observed in popular children's literature? RQ3 How can the results be used for semantic-based DDL focused on emotion words for young EFL learners? We created a specialised corpus comprising 28 popular works of children's literature based on the best-selling works (2017-2020) in the UK, US, and Japan. The following three outcomes were achieved: (1) Word level by Japanese English Textbooks 2020 of the New Word Level Checker was used to determine the appropriateness for young EFL learners. (2) Extracted 100 high-frequency emotion words (50 positive and 50 negative), which were categorised into six semantic categories using the CREL Semantic Analysis System. (3) Accordingly, a semantic-based DDL focused on emotion words given appropriate young EFL learner level is suggested. This study addresses the gap regarding semantic-based DDL for young EFL learners.

Keywords: Children's literature, emotion word, DDL

1. Introduction

Reading children's literature and fiction is considered to play an important role in developing pre-school and primary-school-age children's feelings. Given authors' intentions when writing children's literature, each word should be carefully chosen, and authors likely give extensive consideration to the meaning and sense of each word, sometimes using the phonetic sound to inspire children's imagination. Thus, an author's chosen words or expressions may be described as intentionally selected for children as readers. This study examines emotion words in popular children's books and considers pedagogical applications as semantic-based data driven learning (DDL) focused on emotion words for young learners of English as a Foreign Language (EFL). However, few studies focus on the emotional expressions in children's literature or the implementation of DDL for young EFL learners in the classroom.

This study identifies the word level and proportion of word usage of frequently used positive- and negative emotion words in a sample of popular children's literature. Based on the results, this study provides suggestions for a semantic-based DDL focused on emotion words.

2. Previous Studies

Previous children's literature studies adopt a variety of perspectives. In terms of the importance and effectiveness of reading literature, some researchers advocate for the benefits from diverse perspectives as follows: parents as a good reading model for children to develop their literacy (Saracho & Spodek, 2010), and the benefits of storybook reading for children to develop their literacy and language skills (Kucirkova and Grover, 2024; Nordström et al. 2024; Cheetham, 2015; Sénéchal et al., 1996; Robbins and Ehri, 1994).

Regarding trends in the selection of children's books for primary school children, Reuter and Druin (2004) investigate how primary school students search for and select books using a digital library (The International Children's Digital Library) and report category differences according to gender and age. In a study on the trends observed in parents' selection of books, Saracho and Spodek (2010) conducted an interview survey with 125 parents of children from five kindergarten classes. Their findings revealed a tendency for parents to select books that reflect contemporary trends, with almost half of the books being modern or fantasy and frequently featuring animal characters.

As a different perspective, Cremin (2008) indicates that teachers lacked sufficient professional knowledge of children's literature, including authors, poets, and creators of illustrated fiction, and identified this as a potential factor affecting young learners' reading development. Focusing on the vocabulary size of children's books, Webb and Macalister (2013) highlight that the vocabulary requirements in children's literature are comparable to those of texts written for older readers. Macalister and Webb (2019) report on the potential for using L1 children's literature in the EFL classroom. Additional research has revealed the vocabulary level of children's literature used a rare vocabulary for native English speakers is too difficult for intermediate English as a Second Language (ESL) students (MacQuillan, 2019; Cheetham, 2015). Senthok Singh and Ang (2019) identified characteristics of lexical bundles usage in the creation of a 1.7-million-word corpus of 30 popular children's novels. They found that prepositional and verb phrases are used much more frequently than noun phrases and other grammatical structures.

Regarding the positive influence of reading children's literature, Ruiqi and Dangui (2023) emphasise the significance of children's five sensory experiences while Nikolaieva and Kravets (2020) highlight the unique

capacity of fiction to foster human sentiment and spiritual values. Moreover, Gardner (2008) utilise 14 collections of children's reading materials and highlighted the usefulness of real-world data, in this case, actual words and word repetitions in authentic reading materials. Furthermore, Cheetham (2015, p. 15) notes that 'Reading children's literature is, in short, good training for reading adult literature. It is also good training for understanding other discourse types which employ the same stylistic and rhetorical techniques they will have learned through reading children's literature'. With a particular focus on the semantic field of emotion, Zen (2018) focus on emotional expressions in children's literature to investigate the characteristic language of such literature. This field covers a range of words, classified as follows: verbs (e.g. cried, laughed), adjectives (e.g. funny, glad), and adverbs (e.g. angrily, anxiously). Their findings indicate that children's literature employs a significantly higher frequency of positive emotional expressions (e.g. gently, joy) and a markedly lower frequency of negative emotional expressions (e.g. angry, frightened) within the corpus.

Previous domestic and overseas corpus studies have examined DDL. Luo and Zhou (2017) examine DDL research on L2 writing from 2010 to 2016, and identified three main streams of DDL studies: (1) the effects of DDL on writing by comparing writing using corpora and dictionaries as resources; (2) the effect of using different types of corpora on writing by comparing general and specialised and native speakers' and learner corpora; and (3) the effectiveness of using Google as a reference tool, which is more flexible than most concorodancers and user friendly.

Some researchers have highlighted problems with implementing DDL in secondary-school classrooms as follows: (1) most studies focus on university-level learners from the upper-intermediate to the advanced level (Klomkaew and Boontam, 2023; Kongcharoen and Thummanuruk, 2023; Hajimia et al., 2022; Chitez and Bercuci, 2019; McGuire, 2019; Kogan et al., 2018; Larsen-Walker, 2017; Nugraha et al., 2016; Boulton, 2015; Chujo et al., 2012). (2) Challenges exist in the educational environment regarding Internet access to use corpora and the lack of enough computers in schools (Boulton, 2009). (3) Both learners' and teachers' lack of knowledge for using corpora (Crosthwaite, 2019; Di Vito, 2019; Hirata, 2019; Schaeffer-Lacroix, 2019; O'Keeffe and Farr, 2003; Johns, 1991; Boulton, 2010). (4) A large gap exists among young and lower-level learners regarding use of general corpora such as the BNC and COCA (Moon and Oh, 2018; Lenko-Szymanska and Boulton, 2015; Granger and Gilquin, 2010). (5) Using the DDL approach is time-consuming, and teachers have limited time to develop DDL materials (Schaeffer-Lacroix, 2019).

Previous studies on DDL for beginner-level university students report the efficiency of computer-based DDL in blended learning, which combines classroom, online, and off-campus work (Chujo et al., 2013a), and the development and efficiency of freeware parallel corpus such as 'AntPConc' (Chujo et al., 2013b).

Regarding the implementation of computer-based DDL at secondary school level, Nishigaki et al. (2022) developed a hybrid DDL application and verified its appropriateness in terms of the proficiency level of Japanese secondary school learners. The feature of collecting the corpus data is considered young EFL learners' level, which mainly consists of the following three resources: 1) school textbooks in Japan, China, Korea, and the U.S., 2) graded readers, 3) EFL materials on the Internet. Regarding the efficiency of paper-based DDL, Nishigaki et al. (2025) prove that paper-based and a hybrid of paper+web-based DDL are effective for grammar instruction at the beginner level in secondary schools. They highlight the benefit of paper-based DDL given limitations in educational environments in terms of using IT tools and suggest avoiding spending too much time on DDL to reduce the learners' burden. They also emphasize editing and simplifying concordance lines and teachers guiding students regarding word usage and grammatical rules.

The above literature review reveals that few studies focus on the emotional expressions in children's literature or the implementation of DDL for young EFL learners in the classroom. Thus, the present study focuses on emotion words in popular children's literature and suggests implementing semantic-based DDL focused on emotions for young EFL learners.

3. Research Design

To identify popular children's literature read by children, particularly at the primary and secondary school levels, amongst both young L1 and L2 learners, the UK, US, and Japan were selected for data collection.

Our linguistic analysis focuses on the semantic field of emotion words in children's literature. The aim is to clarify how authors employ emotion words; in other words, what kind of emotion words do children encounter when reading children's literature? Based on the analysis, the present study suggests an online semantic-based DDL focused on emotion in popular children's literature.

3.1 Research questions

Based on the research design, the present study set the following research questions:

RQ1: What word levels do authors use in popular children's literature?

RQ2: What high frequency emotion words are observed in popular children's literature?

RQ3: How can the results be used for semantic-based DDL focused on emotion words for young EFL learners?

3.2 Data collection and building corpora

The data consist of 28 popular children's books (17 bestselling books from a bookstore in the UK, 6 from a bookstore in Japan, and 5 from the New York Times in the US) from 2017 to 2020. The collection and building of the corpus started in 2017 but required considerable time due to the COVID-19 pandemic, when research was halted as daily classes needed to be adapted to a remote format.

The aim is to create a specialised corpus based on the popular children's literature among children of the same generation in different countries. Each of the 28 works in the corpora was assigned a code from C1 to C28. The list of the works used to build corpus data for this research are presented in Table 1.

A UK bookshop	
Corpus	Book Title
C1	Amelia Fang and the Barbaric Ball
C2	Beast Quest
C3	Harry Potter and the Philosopher's Stone
C4	My Naughty Little Sister
C5	George's Marvellous Medicine
C6	Fantastic Mr Fox
C7	The Twits
C8	The Boy at the Back of the Class
C9	The Creakers
C10	The Enchanted Wood
C11	The Fox and the Ghost King
C12	The Impossible Boy
C13	The Owl Who Was Afraid of the Dark
C14	The Star Outside My Window
C15	The Unwanted Puppy
C16	The Worst Witch
C17	Toto the Ninja Cat and the Great Snake Escape

A Japanese bookshop	
Corpus	Book Title
C18	Winnie the Pooh
C19	Mary Poppins
C20	Holes
C21	The Little Prince
C22	The Crimes of Grindelwald
C23	Alice's Adventures in Wonderland
The New York Times	
Corpus	Book Title
C24	Look Both Ways
C25	Sweeping Up the Heart
C26	All the Greys on Greene Street
C27	Some Places More Than Others
C28	My Jasper June

Table 1. Study sample of popular children's literature

The following procedures were used to construct the children's literature corpus: 1) Textual data were scanned and manually checked because some English characters and signs are garbled, which is a known problem for corpus analyses using a computer, 2) After carefully reviewing the text data, this study constructed a corpus consisting of 28 works of children's literature. 3) Next, using the Wmatrix 5 corpus analysis tool (Rayson, 2022), we conducted a cross-tabulation analysis to calculate the TTR (type-to-token ratio).

Figure 1 shows the TTR (type-to-token ratio) percentages for each of the 28 works, sorted in descending order of TTR value.

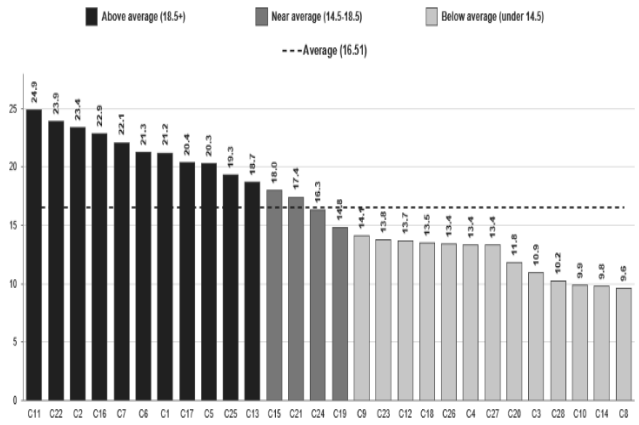

Figure 1. Type/Token Ratio of each work in the Corpus

The highest TTR is 24.90 in C11 (The fox and the Ghost King), while the lowest is 9.64 in C8 (The boy at the Back of the Class).

The number of tokens, types, and type/token ratio are reported in Table 2. The total number of words in the 28 works of children's literature is 828,290 and the average TTR is 16.51.

	C11	C22	C2	C16	C7	C6	C1	C17	C5	C25
Number of tokens	5515	18871	9701	9126	8064	9388	15824	15284	11930	21159
Number of types	1373	4513	2266	2086	1783	1996	3359	3112	2427	4087
Type–token ratio (%)	24.90	23.92	23.36	22.86	22.11	21.26	21.23	20.36	20.34	19.32
	C13	**C15**	**C21**	**C24**	**C19**	**C9**	**C23**	**C12**	**C18**	**C26**
Number of tokens	9866	12029	14205	33114	35754	41861	25427	48655	20774	69684
Number of types	1843	2165	2476	5396	5296	5904	3508	6653	2806	9364
Type–token ratio (%)	18.68	18.00	17.43	16.30	14.81	14.10	13.80	13.67	13.51	13.44
	C4	**C27**	**C20**	**C3**	**C28**	**C10**	**C14**	**C8**	**Average**	
Number of tokens	16953	34737	45056	76286	64252	40503	58009	56263		
Number of types	2265	4641	5319	8335	6552	4013	5674	5421		
Type–token ratio (%)	13.36	13.36	11.81	10.93	10.20	9.91	9.78	9.64	**16.51**	

Table 2. Token Counts, Type Counts, and Type–Token Ratios by Corpus

Note. Corpora are ordered by type–token ratio (TTR); the overall mean TTR is 16.51%.

4. Method

First, to determine the learner level for each work, this study used the New Word Level Checker (Mizumoto, 2022). Second, we identified the top 50 positive and negative emotion words in the 28 works to develop semantic-based DDL using the Wmatrix 5 corpus analysis tool (Rayson, 2022). Based on the results, this study suggests a semantic-based DDL approach in the classroom.

5. Data Analysis and Discussion

RQ1 What word levels do the authors use in popular children's literature?

This study identifies the word levels in a sample of current popular children's books, which enables us to grasp whether the word levels in the DDL concordance lines are appropriate for young EFL learners. If the concordance lines contain more advanced vocabulary, learners may struggle to determine the keywords. Regarding word levels, this study adopted JET 2020 (Japanese English Textbooks 2020) of the New Word Level Checker developed by Mizumoto (2022) to determine the appropriateness for young EFL learners. The JET 2020 word level list consists of the following six main levels: PropNoun_Num (proper noun and numeral as a word expressing a number with which learners are already familiar), elementary (628 words), junior high (1,758 words), which is further subdivided into three levels (JH1, JH2, and JH3: first-, second-, and third-year junior high school, respectively), and SH (high school) (3,039 words); the total number of words is 5,425.

The results of the coverage rate of each word level in the 28 works based on JET 2020 of the New Word Level Checker are reported in Figure 2.

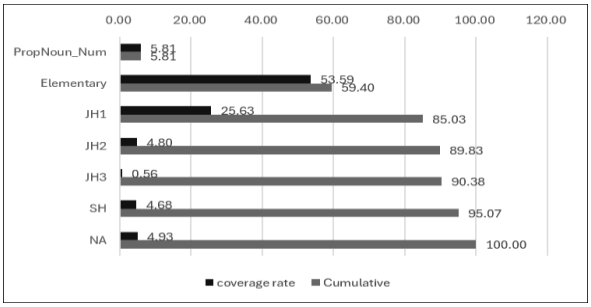

Figure 2 Coverage rates for each word level

Note. Word levels: PropNoun_Num: proper noun and numeral with which learners are already familiar, Elementary: elementary. JH1, JH2, JH3: first-, second-, and third-year junior high, respectively. SH: senior high. NA: Not applicable. PropNoun_Num and Elementary are treated as basic vocabulary levels.

At the basic level, the coverage rate of the PropNoun_Num and Elementary is 5.81% and 53.59%, respectively. The cumulative coverage rate of the two-word level is 59.40%; in other words, approximately 60% of the sample works include the two-word basic level vocabulary.

The actual coverage rate for JH1, JH2, and JH3 are 25.63%, 4.80%, and 0.56%, respectively. The cumulative coverage rate of the JH 1, JH 2, and JH 3 are 85.03%, 89.83%, and 90.38%, respectively. The results indicate that a high proportion of the authors use a basic word level in their works. The actual coverage and cumulative coverage rates for SH are 4.68% and 95.7%, respectively, which indicates a suitable readability level for junior high school learners.

RQ2 What high-frequency emotion words are observed in popular children's literature?

The usage of emotion words in children's literature plays an important role for both authors to describe character's feelings and readers to imagine the situation through such feelings. To grasp high-frequency emotion words in the literature sample, this study first identifies the Top 100 high-frequency emotion words using 'emotion word + USAS tag'. Interestingly, the coverage rates of the positive and negative emotion words are 51 and 49 (i.e., the number of occurrences observed among the top 100 emotion words), respectively. In other words, the ratio is approximately fifty-fifty. Based on the results, this study identifies the usage ratio of each emotion word by the UCREL Semantic Analysis System (USAS) tag classified into six semantic categories. In addition, 50 positive- and negative- emotion words were extracted to create a recommended word list for teachers to conduct semantic-based DDL in the classroom.

5.1 Semantic categories of emotion words

The six semantic categories of Emotion words based on the USAS tag are presented in Table 3.

Sem. tag	Category labels	Explanation
E1	General	General terms depicting emotional actions, states and processes
E2	Liking	Terms depicting fondness/affection/partiality/attachment, or the lack of
E3	Calm/Violent/Angry	Terms depicting (level of) serenity/composure/anger/violence
E4	Happy/sad	Entries are sub-classified into the following:
E4.1	Happy/sad: Happy	Terms depicting (level of) happiness
E4.2	Happy/sad: Contentment	Terms depicting (level of) contentment
E5	Fear/bravery/shock	Terms relating to (level of) trepidation/courage/surprise, etc.
E6	Worry, concern, confident	Terms relating to (level of) apprehension/confidence, etc.

Table 3. Semantic categories of emotion words categorised by USAS tag

The semantic category E4 *(Happy/sad)* has two sub-categories: E4.1 *(Happy/Sad: Happy)* and E4.2 *(Happy/sad: Contentment)*. Based on the list, we extracted the top 50 high frequency emotion words using the Wmatrix 5 corpus analysis tool.

5.2 Coverage rate of semantic categories: Positive emotion words

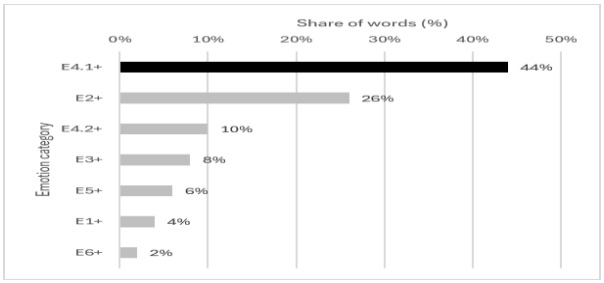

Figure 3. Coverage rate of semantic categories of positive emotion words

The coverage rates of semantic categories are presented in Figure 3. E4.1+ *(Happy/sad: Happy)* is the highest frequency at 22 (44%), followed by E2 *(Liking)* with 13 (26%). The coverage of the two high frequency emotional categories (E4.1+ and E2+) are much more clearly than the following the six emotional categories: E4.2+ *(Happy/sad: Contentment)* = 5 (10%), E3+ *(Calm/Violent/Angry)* = 4 (8%), E5+ *(Fear/bravery/shock)* = 3 (6%), E1+ = 2 (4%), and E6+ = 1 (2) %.

5.3 Top 50 positive emotion words

In addition, extracting and identifying the positive emotion words is valuable to implement semantic-based DDL. The top 50 positive emotion words categorized by USAS tag are listed in Table 4. The extracted words are presented in an orthographic form considering the learning stage of the main grammar structures in the syllabus of Japanese textbooks of English in junior high school from the first- to the third-year level. In semantic-based DDL, the difference in the orthographic form is considered a

difficulty for young EFL learners when guessing the keyword in context in concordance lines.

For classroom-based semantic DDL, this classification of the top 50 positive emotion words would be a useful teaching material.

USAS Semantic Tag	Semantic Category	Top 50 Positive Emotion Words
E1+	General	tone, feel
E2+	Liking	feel, like, liked, likes, dear, love, loved, darling, enjoy, live, fancy, fan, favourite, favorite
E3+	Calm/Violent/Angry	rest, gently, calm, softly
E4+	Happy/Sad	—
E4.1+	Happy/Sad: Happy	smile, smiled, smiling, smiles, funny, happy, happily, laugh, laughing, laughed, laughter, fun, grin, grinned, grinning, joke, relief, delight, delighted, joy, giggled, beamed
E4.2+	Happy/Sad: Contentment	glad, pleased, proud, pleasure
E5+	Fear/Bravery/Shock	brave, dare, courage
E6+	Worry/Concern/Confidence	mind

Table 4 Top 50 positive emotion words categorized by USAS tag

5.4 Coverage rate of semantic categories: Negative emotion words

In the next stage, we focus on the top 50 negative emotion words used in the study sample and identify the coverage rate of the six semantic categories of these words (see Figure 4).

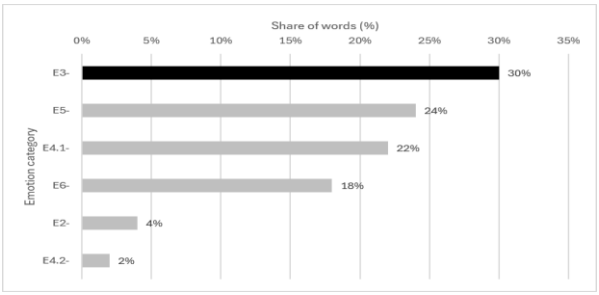

Figure 4. Coverage rate of semantic categories: Top 50 negative emotion words

The semantic category of E1 is not observed for negative emotion words and the following four categories have a relatively high coverage rate: E3- *(Calm/Violent/Angry)* with 15 (30%), E5- *(Fear/Brave/Shock)* with 12 (24%), E4.1- *(Happy/Sad: Happy)* with 11 (22%), and E6- *(Worry concern confident)* with nine (18%). Furthermore, the coverage rate is higher than that of two other negative emotion categories: E2- *(Liking)* with two (4%) and E4.2- *(Happy/Sad: Contentment)* with one (2%). Therefore, the higher coverage rate of the four categories (E3-, E5-, E4.1-E6) than the two other negative semantic categories (E2-, E4.2-) reveals a different trend in the coverage rate distribution compared to the positive emotion words.

5.5 Top 50 negative emotion words

After establishing the distribution of semantic categories of the top 50 negative emotion words, extracting these words would also be beneficial for teachers for use in semantic-based DDL in the classroom. The results for the top 50 high-frequency negative emotion words and their semantic categories based on the USAS tag are presented in Table 5.

USAS Semantic Tag	Semantic Category	Top 50 Negative Emotion Words
E2−	Liking	hate, hated
E3−	Calm/Violent/Angry	angry, hit, angrily, kicked, bullies, poked, annoying, fierce, peeves, annoyed, furious, kick, kicking, anger, crossly
E4	Happy/Sad	—
E4.1−	Happy/Sad: Sadness	cried, sad, cry, crying, upset, sadly, miserable, desperate, sadness, looked, embarrassed
E4.2−	Happy/Sad: Discontentment	disappointed
E5−	Fear/Bravery/Shock	scared, afraid, frightened, fear, shock, scary, horror, looked, panic, terrifying, terrified, alarm
E6−	Worry/Concern/Confidence	worry, trouble, worried, care, nervous, anxiously, nervously, anxious, bother

Table 5. Top 50 negative emotion words categorized by USAS tag

5.6 Semantic-based DDL: A case study

The study results can be used to create semantic-based DDL materials.

5.6.1 Semantic-based DDL for university students via *Moodle*

This case study involved 12 L2 learners (6 third year, 6 fourth year) on the Corpus Linguistics of the Department Communication Course at the University of Shimane in June 2026. The procedures of the semantic DDL are as follows:

1) A PDF file showing the concordance lines focused on the word immediately before and after the keyword (*****) on *Moodle* (an example of part of the concordance of the semantic-based DDL based on the keyword '*sad*' is presented in Figure 5).

shell . There 's **nothing**	*******	about an old shell ... "
he 'd seen all the commercials **of**	*******	pups locked behind cages , sick and
still tied balloons at the ends **of**	*******	ribbons . Two strands of
, that the world is full **of**	*******	people , sick people , people with
, Jasper was sitting on the **one**	*******	kitchen chair , eating peanut butter from
And they always sound angry **or**	*******	when they talk about it .
particularly tough or loud or funny **or**	*******	or weird or even smelly .
them goodbye , and Joe felt **quite**	*******	to see them go .
And since he was feeling **rather**	*******	on account of remembering his own little
do when you 're really , **really**	*******	about something , ' she said .
down . Dad got **really**	*******	, ' continued Ben .
Sad . Very sad. **Unbearably**	*******	. Poor aunt Emily .
Suddenly , he was wearing the **same**	*******	, handsome look on his face I
Mock Turtle in the distance, **sitting**	*******	and lonely on a little ledge of
Squashy . He looked **so**	*******	; he was n't even bouncing or
but then the next you feel **so**	*******	that you want to hide in your
of times hut it made me **so**	*******	... " She held up
tucked down and bis eyes looked **so**	*******	. He was unhappy ,

Figure 5. An example of concordance lines of the semantic DDL distributed as a PDF file on *Moodle*

2) The following three assignment tasks are presented:

Task 1: Write about your any discoveries about the word immediately before the keyword.

Task 2: Write about your any discoveries about the word immediately after the keyword.

Task 3: How did you guess the keyword?

As an individual task, students read the concordance lines as a PDF file. Students then wrote about what they noticed about the word immediately before and after the keyword and submitted this on *Moodle* (Task 1 and 2, respectively). Working in groups of four, they then shared what they noticed about the keyword in context. Finally, they submitted the reasons they expected the keyword on *Moodle* (Task 3).

5.6.2 Students' comments

In general, collecting students' comments after implementing paper-based DDL and giving feedback requires considerable time for teachers. Accordingly, this study develops an online semantic-based DDL administered via *Moodle* to collect students' comments more effectively. In addition, text mining analysis via *KH Coder* is used to visualize and grasp the trends in students' comments. The text mining analysis involves creating text data from the students' comments on *Moodle* regarding the word immediately before and after the keyword in context in the concordance lines. A co-occurrence network analysis is adopted to visualize the relationships between the word immediately before and after the keyword and the keyword itself.

The results of the co-occurrence network analysis are presented in Figure 6.

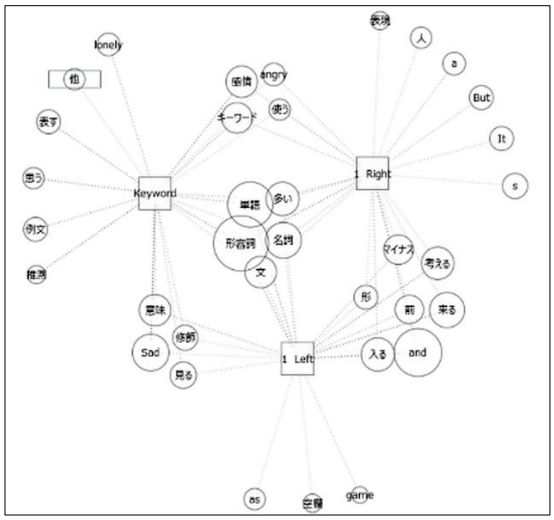

Figure 6. Co-occurrence network of students' comments in the semantic-based DDL

The co-occurrence network analysis reveals students' conceptions regarding the word immediately before and after the keyword in the context of the concordance lines. Students choices regarding the keyword are also extracted. Regarding the word immediately before the keyword, students predicted the following: '*omou*' (think), '*kuru*' (come), '*hairu*' (fill), '*miru*' (see) and '*syuushoku*' (*modify)*. These words suggest that the learner tries to identify some features of the language usage when trying to determine the keyword. Regarding the word immediately after the keyword, the following words were noted: '*tsukau*' (use), '*ki-wa-do*' (keyword), '*kanjyou*' (feeling), 'shikasi' (But)', 'sorewa' (It), '*hito*' (person), '*hyougen*' (expression). The students' observations suggested that they noted the different subjects at the beginning of a new sentence after the keyword. Regarding the process of identifying the keyword, the following words were noted: 'suisoku' (expect), 'omou' (think), 'kanjiru' (feel), 'arawasu (express), and 'tekisetsu' (appropriate). The following words related to learners' discoveries were also observed: '*ooi*' (many), '*keiyousi*' (adjective), '*bun*' (sentence), '*meisi*' (noun), and '*tango*' (word). The words indicate that learners attempted to establish some grammar rules regarding the part of speech of the keyword in a sentence.

5.7 Semantic-based DDL in the classroom for young EFL learners

Previous studies report the implementation of paper- and computer-based DDL, as well as online DDL tools for learners at an elementary and a junior high school. Teachers can conduct paper-based DDL in the classroom, which can overcome lack of computers and/or insufficient Wi-Fi networks. However, such approach requires preparation to print the DDL materials and is time consuming. On the other hand, online DDL tools, which focus on learning main grammar structures, are designed to be user friendly for both teachers and beginner-level learners.

This study suggests semantic-based DDL focused on positive and negative emotion words derived from corpus data based on 28 popular works of children's literature. This study also suggests using the semantic-based DDL as a complement to IT tools such as Moodle, Teams, and Google Classroom in secondary education.

However, given the educational circumstances, as a per English-lesson time is 50 minutes in Japanese junior- and senior high school, the semantic-based DDL focused on emotion words could serve as a bridge between learning in the classroom and at home as self-study based on learner's motivation to read current popular children's literature.

6. Conclusion

Based on a sample of 28 current popular children's books from bestselling lists in the UK, the US, and Japan from 2017–2020, this study investigated the use of emotion words and constructed a corpus for online semantic-based DDL. Analysis of the word levels by the New Word Level Checker, JET 2020 indicates that the coverage ratio of the word level is more than 80% of the first- to third-year level, which is an appropriate level for reading the concordance lines of the semantic-based DDL. The top 50 positive and negative emotion words as a recommended list for teachers and young EFL learners to learn emotion words through the semantic-based DDL could be a valuable resource for effectively teaching and learning such words in popular children's literature. Based on the results, this study suggests a semantic-based DDL in the classroom using IT tools like *Moodle*, *Teams*, and *Google Classroom*. In addition, teachers can easily share the assignment tasks regarding keywords in context.

The semantic-based DDL focused on emotion words could provide teachers and young EFL learners a new approach for learning emotion words based on current popular children's books.

7. Limitations

Based on the bestselling lists of a UK bookstore, The New York Times, and a Japanese bookstore from 2017 to 2020, this study identified emotion words in 28 current popular children books and created a specialised corpus. However, the number of books in the corpus is small due to the time required for data collection and annotation. A large corpus of children's literature, the Oxford Children's Corpus (OCC) is available, but the use of the data is limited for researchers. In addition, some corpus data seem to be out-of-date and some were collected from parts of chapters rather than complete children's books. As this study attempted to implement as a case study for university students, further research should focus on young EFL learners from the second- or third-year level of junior to senior high school. This study may be helpful for teachers and young EFL learners to experience semantic-based DDL to learn emotion words and grow as autonomous learners.

However, when undertaking this study, it is important to consider the copyright limitations when using original children's literature. According to copyright rules, as the creator of this material, I can use it directly in classroom teaching. However, I cannot provide other teachers with the same materials. Therefore, teachers are encouraged to use these techniques and implement their own original materials in their classrooms.

8. Acknowledgements

I would like to express my appreciation to P. Rayson for providing the Wmatrix 5 corpus analysis tool and helpful suggestions. I thank students taking the Corpus Linguistics course at the University of Shimane for their cooperation and participation as a case study for my research.

9. References

9.1 Bibliographical References

Boulton, A. (2009). Data-driven learning: On paper, in practice. *Corpus Linguistics in Language Teaching*, 17–52.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534–572.

Boulton, A. (2015). Applying data-driven learning to the web. *Multiple Affordances of Language Corpora for Data-Driven Learning*, 267–295.

Cheetham, D. (2015). Extensive reading of children's literature in first, second, and foreign language vocabulary acquisition. *CLELE Journal*, 3(2), 1–23.

Chitez, M. and Bercuci, L. (2019). Data-driven learning in ESP university settings in Romania: Multiple corpus consultation approaches for academic writing support. *CALL and Complexity–Short papers from EUROCALL*, 2019, 75–81.

Chujo, K., Oghigian, K., and Nishigaki, C. (2012). Beginner level EFL DDL using a parallel web-based concordancer. In Pan-Asian Conference on Language Teaching and Learning (PAC) 2012 and 9th FEELTA International Conference, Far Eastern Federal University, Vladivostok, Russia (pp. 1–5).

Chujo., K. Oghigian, K., and Uchibori, A. (2013a) Comparing computer-based and paper-based DDL in the beginner level English classroom. *Journal of the College of Industrial Technology*, Nihon University. 1–11.

Chujo, K., Anthony, L., Oghigian, K., and Yokota, K. (2013b). Teaching remedial grammar through data-driven learning using AntPConc. *Taiwan International ESP Journal*, Vol. 5. 65–90.

Cremin, T., Mottram, M., Bearne, E., and Goodwin, P. (2008). Exploring teachers' knowledge of children's literature. *Cambridge Journal of Education*, 38(4), 449–464.

Crosthwaite, P. (Ed.). (2019). Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners. Routledge.

Di Vito, S. (2019). Teaching French to young learners through DDL. In P. Crosthwaite (Ed.), Data-driven learning for the next generation (pp. 171–186). Routledge.

Gardner, D. (2008). Vocabulary recycling in children's authentic reading materials: A corpus-based investigation of narrow reading. Reading in a foreign language. 20-1, 92-122.

Gilquin, G. and Granger, S. (2010). How can data-driven learning be used in language teaching? In The Routledge handbook of corpus linguistics (pp. 359–370). Routledge.

Hajimia, H., Sarjit Singh, M.K., Mohd Radzi, N.A., Mohd Nordin, N.R., Che Ismail, N.A., and Yousef Aljarrah, M.M. (2022). Use of corpus-integrated data-driven learning approach in the language classroom: Malaysian educator's perspectives. Res Militaris.

Hirata, E. (2019). The development of a multimodal corpus tool for young EFL learners: A case study on the integration of DDL in teacher education. In P. Crosthwaite (Ed.), Data-driven learning for the next generation (pp. 88–105). Routledge.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials (Vol. 4, pp. 1–16). na.

Klomkaew, T. and Boontam, P. (2023). Effectiveness of grammar learning through data-driven learning (DDL) with undergraduate students. *Shanlax International Journal of Education*, 11(4), 30–44.

Kogan, M., Yaroshevich, A., and Ni, O. (2018). Corpus-based teaching of German compound nouns and lexical bundles for improving academic writing skills. *Lidil. Revue de linguistique et de didactique des langues*, (58).

Kongcharoen, P.A. and Thummanuruk, W. (2023). A corpus-based study of the synonyms: Perfect, flawless, impeccable. *THAITESOL Journal*, 36(1), 1–20.

Kucirkova, N. and Grøver, V. (2024). The importance of embodiment and agency in parents' positive attitudes

towards shared reading with their children. *Early Childhood Education Journal*, 52(1), 221–230.

Larsen-Walker, M. (2017). Can data driven learning address L2 writers' habitual errors with English linking adverbials? *System*, 69, 26–37.

Leńko-Szymańska, A. and Boulton, A. (2015). Introduction: Data-driven learning in language pedagogy. In Multiple affordances of language corpora for data-driven learning (pp. 1–14). John Benjamins Publishing Company.

Luo, Q. and Zhou, J. (2017). Data-driven learning in second language writing class: A survey of empirical Studies. International *Journal of Emerging Technologies in Learning*, 12(3).

Macalister, J. (2019). Can L1 children's literature be used in the English language classroom? High frequency words in writing for children. Reading in a foreign language.

McQuillan, J. (2019). Is children's literature as hard as scholarly articles about children's literature? A comment on Macalister and Webb (2019). *Reading in a Foreign Language*, 31(2), 302–304.

Mizumoto, A. (2022). New Word Level Checker の概要. メソドロジー研究部会報告論集, 12, 1–24.

Moon, S. and Oh, S.Y. (2018). Unlearning overgenerated be through data-driven learning in the secondary EFL classroom. *ReCALL*, 30(1), 48–67.

Mariia, N., & Olena, K. (2020). Education of Primary School Pupils of Humanistic Values by Means of Children's Literature. Теорія і практика навчання іноземних мов.

Nishigaki, C., Akasegawa, S. and Oghigian, K. (2022). Development of an online DDL tool for secondary school learners. *Bulletin of the Faculty of Education*, Chiba University, 70, 289–297.

Nishigaki. C., et al. (2025). Optimizing the efficiency of L2 grammar learning in secondary schools: Paper-based DDL and Paper+ Web-based DDL. *Bulletin of the Faculty of Education*, Chiba University, 73, 389–396.

Nordström, A., Kumpulainen, K. and Rajala, A. (2024). Unfolding joy in young children's literacy practices in a Finnish early years classroom. *Journal of Early Childhood Literacy*, 24(1), 3–19.

Nugraha, S.I., Miftakh, F., and Wachyudi, K. (2016, November). Teaching grammar through data-driven learning (DDL) approach. In Ninth International Conference on Applied Linguistics (CONAPLIN 9) (pp. 300–303). Atlantis Press.

O'Keefe, A. and Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly*, 37(3), 389–418.

Rayson, P. (2009) Wmatrix: a web-based corpus processing environment, Computing Department, Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/

Reuter, K. and Druin, A. (2004). Bringing together children and books: An initial descriptive study of children's book searching and selection behavior in a digital library. Proceedings of the American Society for Information Science and Technology, 41(1), 339–348.

Robbins, C. and Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology*, 86(1), 54.

Ruiqi, P. and Dangui, X. (2023). The significance of teaching children's picture books with. *Applied & Educational Psychology*, 4(11), 110–114.

Saracho, O.N. and Spodek, B. (2010). Families' selection of children's literature books. *Early Childhood Education Journal*, 37, 401–409.

Schaeffer-Lacroix, E. (2019). Barriers to trainee teachers' corpus use. In Data-driven learning for the next generation (pp. 47–64). Routledge.

Sénéchal, M., LeFevre, J.A., Hudson, E., and Lawson, E.P. (1996). Knowledge of storybooks as a predictor of young children's vocabulary. *Journal of Educational Psychology*, 88(3), 520.

Senthok Singh, A.K. and Ang, L.H. (2019). Lexical bundles in selected children's fiction: A corpus-based analysis. *Journal of Creative Practices in Language Learning and Teaching (CPLT)*, 7(2), 12–28.

Webb, S. and Macalister, J. (2013). Is text written for children useful for L2 extensive reading? *TESOL Quarterly*, 47(2), 300–322.

Zen, E.L. (2018). Corpus-driven analysis on the language of children's literature. In Proceedings of the 1st International Conference on Recent Innovations.

9.2 Language Resource References

Anderson, L.E. (2017). Amelia fang and the barbaric ball. Egmont.

Blyton, E. (2014). The Enchanted wood. Egmont.

Brade, A. (2007). Beast quest, Ferno the fire dragon. Orchard.

Brooks, B. (2019). The impossible boy. Quercus.

Carrol, L. (2015). Alice's adventures wonderland. Macmillan.

Dahl, R. (2016). Fantastic Mr fox. Puffin Books.

Dahl, R. (2016). George's marvellous medicine. Puffin.

Dalph, R. (2016). The twits. Puffin.

Fletcher, T. (2017). The creakers. Puffin.

Henkes, K. (2020). Sweeping up the heart. Greenwillow.

Hughes, S. (2017). My naughty little sister. Egmont.

Milne, A.A. (1992). Winnie-the-pooh. Puffin.

Morpurgo, M. (2016). The fox and the ghost king. Harper Collins.

Muppy, J. (2013). The worst witch. Puffin.

O'Leary, D. (2017). Toto ninja cat and the great snake escape. Hodder.

Rauf, Q.O. (2018). The boy at the back of the class. Orion.

Rauf, Q.O. (2019) The star outside my window. Orion.

Reynolds, J. (2019). Look both ways. A Caitlym Dlouhy.

Rowling, J.K. (1997). Harry Potter and the philosopher's stone. Bloomsbury.

Rowling, J.K. (2019). The crimes of Grindelwald. Sphere.

Sachar, L. (2015). Holes. Yearling.

Saint Exupery, A.D. (2000). The little prince. Mariner.

Snyder, L. (2019). My jasper June. Walden Pond.

Tomlinson, J. (2014). The owl who was afraid of the dark. Egmont.

Travers, P.L. (1997). Mary Poppins. Harcourt.

Tucker, L. (2019). All the greys on Greene Street. Puffin.

Watson, R. (2019). Some places more than others. Bloomsbury.

Webb, H. (2018). The unwanted puppy. Stripes.