My girl, old girl: Keywords, Collocations and Gender in British Children's Fiction

Elizabeth Poynter
e.poynter@leedsbeckett.ac.uk
Carnegie School of Education, Leeds Beckett University, UNITED KINGDOM

Abstract

The importance of children's literature in the development of gender roles and gender identity is widely acknowledged. In the mid twentieth century books were the primary medium for most children, and the popular British fiction of the period is often considered to present strong gender stereotypes, despite the enormous social changes taking place (World War II, the Women's Movement). This study builds on previous content analysis and small-scale linguistic analysis to explore the extent of such stereotyping. It uses two digitised corpora, one of adventure books aimed at girls and boys, and the other of books for girls, from the period 1930-1970. With the corpus linguistics software LancsBox it examines keywords of each corpus against the other, and further against a contemporary (1961) corpus of general English prose. It then looks at collocations of GIRL, BOY, WOMAN, MAN. While there are some distinctions between the two genres, there are also considerable similarities. Some gender stereotyping occurs in use of reporting verbs and physical description, but there is no clear polarisation of gender roles in the collocations, and females are referenced more frequently in the adventure texts than in general prose of the same period. This tends to support previous findings, that the popular children's literature of this period is more nuanced than is often assumed.

Introduction

Children's literature is widely believed to have an influential role in socialisation (Hunt, 2011; Tsao, 2020), including the development of gender identity. Nodelmann (2008) says that one of the main functions of children's literature is to teach them the meaning of being a girl or a boy, and Zanfabro (2017, p. 5) points out that '[g]ender is performatively constructed' in such literature in a range of ways: in the characters, the language, the storyline and indeed the book covers. Evidence suggests that children as young as pre-school may understand gender stereotypes (Skočajić et al., 2020), and exposure to gender stereotypes in the media generally may affect

Elizabeth Poynter

children in a variety of ways, such as occupational choice (Gehrau et al., 2016).

It appears that even in the twenty-first century, children's literature presents gender stereotypes. Anderson et al. (2023) noted significant disparities in the portrayals of mothers and fathers respectively (in picture books), and Tsao (2020), surveying a number of studies, concurred that gender bias was still very common in children's books.

As Sunderland (2011) says, it is important to have a linguistic dimension to any analysis of texts. Corpus linguistics provides a tool for this. For example, Hunt (2011) used corpus linguistics to compare the Narnia books from the 1950s with the Harry Potter series, and more recently (2019) she has extended this to the seven best-selling series aimed at eleven-year-olds in the past century, which included the two aforementioned. Looking at collocations around body parts through the lens of gender, she found a somewhat depressing lack of progress: for instance, male eyes in the texts are more likely to express anger, female eyes laughter or tears, and this did not change significantly over time. On the other hand, Cermáková and Mahlberg (2021) found that, while contemporary children's literature still demonstrates a marked imbalance in the representation of female and male characters, a comparison with a nineteenth-century corpus showed subtle changes, possibly indicating an increase in female independence and confidence.

In order to assess progress (if any) in gender representation in modern children's literature it is useful to know how such literature represented gender in the past. The period of British children's literature from the 1930s to the 1960s has frequently been dismissed by the critics (Eyre, 1972; Rowe Townsend, 1990), or criticised as sexist (and racist and middle-class). Yet a number of authors, such as Enid Blyton, Captain Johns ('Biggles') and Elinor Brent-Dyer were enormously popular and are still in print. (As Rudd (2009) reports, Blyton sells around eleven million volumes a year, even outselling J. K. Rowling.) Early dismissals of this period in children's literature are not frequently challenged, although Rudd (2000; 2009) has

looked critically at Blyton and Butts (2018; 2021) at Captain Johns' hero Biggles. The authors publishing in this period were mostly born at the end of the nineteenth century, but they were writing in a time of considerable social change, including World War Two and culminating in Second Wave Feminism in the 1960s. How did their books influence their child readers; did they reinforce traditional stereotypes, as they are generally accused of doing, or did they in any way challenge these? Adventure fiction in particular offers the opportunity to explore this question: beginning around 1930 there arose a genre of 'mixed' adventure stories, usually featuring a group of girls and boys and aimed at both female and male readers.

A previous exploration of a core of one hundred and twenty-six adventure books published between 1930 and 1970 (Poynter, 2018), largely through content analysis, found that, whilst they in some ways reflected the gender stereotypes prevalent at this period, they also differed in sometimes quite significant ways. They presented girls with a range of active and authoritative role models, and to a lesser extent offered a range of masculinities. Corpus linguistics provides a quantitative means of triangulating these qualitative findings. A small corpus of eight texts manually examined gave little indication of gender stereotyping as regards character descriptors (Poynter, 2020), but this study was clearly limited in scope.

Two features which can be explored through corpus linguistics are keywords and collocations. The goal of a keyword analysis is to identify significant lexical differences between two corpora (Baker, 2004). Keywords in themselves do not constitute discourse, but they may 'direct the researcher to important concepts in a text (in relation to other texts) that may help to highlight the existence of types of (embedded) discourse or ideology' (Baker, 2004, p.348). Koller and Mautner (2004) argue that '...recurrent discursive phenomena that are revealed in large corpora in the form of keywords and collocations offer an observable record of the unconscious behaviours through which dominant meanings are discursively reproduced' (p.217). A keyword analysis may

highlight stereotypes within a text or corpus of texts, which may indeed be below the level of consciousness but are likely nonetheless to be absorbed by the reader, especially the child reader (Hunt, 2015, p.282).

Studies of collocations have also been used to explore the discourse of gender in various ways. Caldas-Coulthard and Moon (2010) looked at *man, woman, boy* and *girl* in the Bank of English corpus and found that 'While men are evaluated in terms of their function and status in society, a woman is evaluated additionally in terms of her appearance and sexuality – even more so in the case of a young woman, whereas young men are evaluated in terms of their behaviour' (p.124). Baker (2014) found a discourse prosody of belligerence around the word *feminist* in the BNC, with collocates such as *militant, axe-grinding* and *outraged*. This discourse prosody (that is, a range of collocates with similar connotations) being expressed over a number of texts is likely to influence the reader, especially if that reader is a child.

By exploring keywords and collocations in two digitised corpora of children's texts from the 1930s-1960s, it is hoped to address the following questions:

1) Does a keyword analysis of a corpus of children's adventure fiction and one of fiction for girls indicate gender stereotyping in either or both of these genres?
2) Do collocations of gender-specific words in these corpora contribute to a discourse prosody of distinct gender roles?

This will provide a basis of comparison for studies of more recent children's literature. With the exception of Hunt's (2015, 2019) work, corpus linguistic studies of mid twentieth-century children's books are lacking in the literature.

Methodology

Two digital corpora were created, one of adventure books (*Adventure*) selected from the larger study (Poynter, 2018), the other of girls' school and family stories (*Girls*) from the same period (see table 1; a full list of authors and texts is in the appendix).

Table 1. Summary of the corpora

Corpus	Description	Dates	Texts	Tokens (i.e. total word-count)
Adventure	children's adventure fiction, 4 female, 3 male authors	1931-69	18	1,232,832
Girls	girls' school and family fiction, 4 female authors	1931-55	8	484,309
LOB	newspapers, fiction, academic writing, general prose	1961	15	1,007,677

These seven authors had been selected as core for the content-analysis study because they were popular rather than highly critically acclaimed (with the exception of Ransome, who was both), prolific, mostly wrote series fiction (to see whether they changed over time), and because several of them also published in other genres, enabling comparisons. Three of the women who wrote adventure fiction also wrote for girls; to increase the size and scope of the *Girls* corpus two works by another well-known writer for girls of the period were added. The two genres are clearly distinguished by their respective blurbs, with marketing targeted at 'boys and girls' and 'girls' respectively. Books for girls are distinct from, though they overlap with, school stories, in that, unlike boys' school stories, even when primarily set in a school, they frequently include family interaction, and are focused on character development rather than 'action'.

This study used LancsBox version 6.0 (Brezina et al., 2020). Keyword analyses were performed comparing the two principal

corpora, *Adventure* and *Girls*. In order to avoid over-emphasis on differences, Baker (2004) suggests using a third corpus to analyse similarities between the target corpora; LOB (Lancaster-Oslo/Bergen) was selected as this reference corpus, because it was available in LancsBox, it is British English and a more relevant date than, for example, the BNC (British National Corpus), although it would have been preferable to use a larger corpus such as the latter. *Adventure* and *Girls* were compared with each other, then with LOB.

The default statistic for calculating keywords in LancsBox is the simple maths parameter (SMP), which as Brezina (2018, p.85) points out makes interpretation of the value of the statistic easier than, for example, using log likelihood. However, this resulted almost entirely in proper nouns, i.e. the names of the characters; log likelihood was therefore selected as a better option (Rayson, 2008). Baker (2004) has suggested that one way to avoid a word appearing key which actually only occurs in a limited number of texts within the corpus is to take this into account and create a cutoff point. For the *Adventure* corpus, taking words from the top seventy keywords list which occur in nine or more of the eighteen texts excludes all the names but one. Because the number of texts was different, the cutoff point for *Girls* could not be exactly the same; in this case selecting words identified as key with a range of at least four out of the eight texts again eliminated most of the names. This keeps the same ratio (50%) of texts, although four is not a wide range. Where keywords threw up questions, such as polysemy, this was further investigated with concordance lines, as outlined in the Results section. (Depending on word frequency, either all available concordance lines or a random sample of one hundred were checked.) Only the top twenty keywords occurring in 50% or more texts are presented and discussed in this paper for reasons of space, and only positive keywords (those which occur with significantly greater frequency) not negative (words which are significantly less frequent).

Collocations were then sought for lemmas GIRL etc. (A lemma is all forms of a word, so for

a verb all conjugations, and for a noun both singular and plural forms.) Further searches were made for the individual words *girl / girls / boy / boys / woman / women / man / men* to check that no distinctions were being missed. After various experiments it proved that a narrow search parameter of 3L0R (i.e. three words to the left of each search term) eliminated most proper names and gave sufficient data to work with. The statistic selected was log dice, which has the advantage of being directly comparable across corpora and brings up exclusive but not only unusual combinations' (Gablasova et al., 2017, p.164). Because these are quite small corpora, a minimum occurrence of three was set for each collocation.

Results and Discussion

The keyword analysis highlighted both genre distinctions and some gender disparities between the two children's corpora. These disparities included reporting verbs. However, when compared with a third, reference corpus, both *Adventure* and *Girls* showed *she* as key, suggesting some prominence given to female characters. Similarly, while the collocates indicated some gender bias in terms of physical appearance, there was also a great deal of similarity in the representations of females and males.

Keywords

Keywords in *Adventure* compared with *Girls* largely fall into two categories: words related to the male gender (masculine pronouns, *man, men, boys* and *boy*) and words related to water-based activities (*island, boat, sea, river* and *water*), as shown in table 2. My earlier content analysis of these and other texts (Poynter, 2018) found that swimming, sailing and rowing were major activities in many adventure texts (and the female protagonists were frequently described as highly competent in these, especially swimming). As the setting for many such books is the school holidays, this is not particularly surprising.

Table 2. Keywords in *Adventure* compared with *Girls*

KW rank	Type	*Adv* freq	Range	*Girls* freq	Range	Log likelihood
1	*he*	16939	18	1877	8	3706.9630
2	*his*	8347	18	1047	8	1588.2189
3	*the*	70622	18	20792	8	1324.1193
4	*bill*	1870	11	???	3	1040.4286
8	*him*	5021	18	729	8	785.4839
9	*captain*	1352	9	20	6	731.6949
14	*man*	1553	18	99	8	524.4787
17	*island*	805	14	8	1	460.2055
18	*boat*	884	16	25	4	416.3087
22	*they*	9632	18	2436	8	394.1078
26	*men*	889	18	39	7	360.6282
37	*we*	6655	18	1695	8	264.5440
40	*sea*	459	15	6	3	253.1737
48	*there*	5633	18	1450	8	214.1665
54	*was*	18360	18	5824	8	193.3912
55	*boys*	663	16	54	7	190.4921
56	*police*	393	16	10	3	190.3652
57	*boy*	579	16	39	7	189.1987
58	*river*	453	12	20	4	183.2806
59	*water*	871	18	101	8	181.5263

Bill being polysemic, concordance lines were examined: of 1,870 occurrences, only twelve (0.64%), in three texts only, related to bills being paid, and there was one reference to a bird's 'razor-like bill'. It just happened that Bill was the name of the protagonist in four of the texts and a more or less major character in four others. The other word for which concordance lines were checked was *captain* (1,352 occurrences; relative frequency 10.966620 per 10,000 words). The overwhelming majority referred to a rank, either in the form of a name or simply 'the captain' (on a ship; examples 1 and 2).

However, there were four examples where the rank was honorary, denoting leadership of the group of child protagonists (examples 3 and 4).

1) *bodies. 'Shan't we go and see the **Captain** of the ship?' he asked. 'Let's go*
2) *fetch the gold?' 'I am going with **Captain** Hanstadt, taking four of his men in*
3) *said Jane. She looked at Oliver. 'Well, **Captain** - what do we do next?' 'I know,*

4) *Peter,' Dickie said quietly. 'We'll try.' 'You're **Captain** of the Club while David's not here,'*

This is a feature of the adventure genre identified in the content analysis of adventure books. It is common for the group of children or teenagers enjoying the adventure to have a (more or less formally) specified leader, who is almost invariably male; even in cases such as M. E. Atkinson's books (see example 3 above), where Jane is explicitly identified as the one who has ideas, her brother Oliver is the leader. Additionally, 17.2% of the names were in fact those of Ransome's two leaders: 'Captain John' and 'Captain Nancy', whose age is never specified but who seem to be about twelve or thirteen. Clearly this is not a real rank, and should therefore probably be regarded as a special example of the usage in sentences 3 and 4. Ransome gives all his child characters a naval rank ('Mate Susan', 'Able-seaman Titty'), and as his books are appreciably longer than those of Blyton and Johns, for instance, and this has not been adjusted for, it seems that this feature of Ransome's books is skewing the data here. Nonetheless, this formalising of leadership within a group of children is not, as

observed, unique to his books. He is however the only author to give (almost) equal leadership to a female character; there are fifty-one references to Captain Nancy (of the Amazon) and eighty-seven to Captain John (of the Swallow), the difference probably due to the fact that there are more chapters on the Swallows than the Amazons; Nancy's role is nevertheless substantial and widely acknowledged by both children and adults.

Table 3. Keywords in *Girls* compared with *Adventure*

KW rank	Type	*Girls* freq	Range	*Adv* freq	Range	Log likelihood
2	*her*	8182	8	9546	18	2619.6121
3	*she*	8961	8	11253	18	2433.1661
4	*school*	914	8	105	18	1717.6178
7	*lucy*	619	4	3	1	1538.2756
10	*matron*	430	6	1	1	1080.1404
11	*miss*	1688	8	1335	18	1022.0921
18	*girls*	932	8	526	17	809.8168
25	*term*	315	8	29	12	621.2802
28	*i*	7130	8	12681	18	590.7239
31	*gillian*	222	4	0	0	564.6092
40	*oh*	770	8	677	18	403.8344
43	*form*	255	8	63	17	373.4086
44	*girl*	625	8	511	18	362.4393
47	*bee*	138	4	1	1	339.7697
50	*pen*	164	6	15	8	323.8830
56	*clare*	104	4	0	0	264.5016
62	*mistress*	132	7	15	9	348.6981
64	*exclaimed*	327	7	221	12	238.0245
67	*herself*	569	8	599	18	222.8727
69	*quite*	683	8	813	18	209.4632

Table 3 shows that words referencing the female gender are key in *Girls*: feminine pronouns and *miss, girls, girl*, plus several proper names; also words related to the topic of school (*school, matron, term, form, mistress*). Note that *bee* largely refers to a character in one text (named Bee; 96% of all occurrences), while *pen* is also a name in three texts (i.e. short for Penelope), with only 12.8% of occurrences being actual objects such as fountain pens which would be part of the 'school life' topic. In the case of *miss*, over 97% of the examples were names of teachers, and a further 1.05% (in five texts by three authors) were a form of address used by a servant or other social inferior to a teenage girl, a marker of the class divisions which are a feature of this period. There were also seven examples of missing someone (i.e. emotion), and nineteen of missing something, such as a train, a meal, a lesson or a chance.

Noteworthy also are *oh* and *exclaimed*, the latter denoting surprise and the former possibly likewise, though *oh* has a range of functions. One would perhaps anticipate that surprise would be more frequent in the adventure genre, but it is possible that girls are more frequently represented as expressing emotion generally. A frequency check of near-synonyms *gasped* and *cried* demonstrated that neither of these was as significantly different between the two genres (table 4). (Relative frequency is per ten thousand tokens.) A further check of the more aggressive *shouted* and *yelled*, suggesting the former is considerably more frequent in *Adventure* than *Girls*, indicates that some, though by no means all, verbs for reporting speech carry a degree

of gender-stereotyping. This is in line with Hunt's (2017) study of speech verbs in popular British children's fiction series over the past seventy years, which found that even in the more modern books there was a strong degree

of gender polarisation; Ruano (2018) has similar findings for reporting verbs in Dickens, although this was of course an earlier period.

Table 4. Relative frequency of 'emotional' reporting verbs

Type	Relative frequency *Girls*	Range (max 8)	Relative frequency *Adventure*	Range (max18)	Log likelihood
exclaimed	6.808004	7	1.792621	12	238.0245
gasped	1.707206	7	1.078817	17	10.2288
cried	2.893922	8	2.295528	16	4.8636
shouted	0.374753	6	2.246859	18	92.4994
yelled	0.333113	3	0.511019	12	2.5341

Hunt's (2017) study of speech verbs found that '[o]verall, the verbs linked to male characters are associated with loudness and power' (p.2), and power was a key theme in my content analysis of children's adventure texts. The keyness of *captain* in *Adventure* supports this; where the word is an actual rank, it invariably applies to male characters, and where it is an honorary rank, as indicating the leader of a group or club, likewise. However, in Ransome we do see that the fictional rank is given to a male in one case (John the captain of the Swallow), and a female in the other (Nancy, the captain of the Amazon).

It appears from the analysis so far that in the 'mixed' adventure genre of the 1930s onwards, like the adventure fiction for boys from which it developed, males were dominant as compared with females. This is not in line with the content analysis, which found largely equal numbers of female and male protagonists, and female characters with notable skills and agency. Could it be the case that the strong female orientation of the girls' books examined here is skewing the data? Both corpora were therefore compared with the LOB corpus of general prose from 1961.

A comparison of *Girls* with LOB identified some keywords which are connected with dialogue (*said, don't, do* and second-person pronoun *you*), some of which also appear as key for *Adventure*. Feminine pronouns and school-life words such as *miss, matron* and *school* continue to be key. However, *Adventure* gives a somewhat different picture. More verbs are key, perhaps reflecting the action focus of

this genre, but the water-based activities have disappeared. It is notable that *she* and *her* have now become key. It is likely that the greater frequency of pronouns is a feature of the narrative genre; narratives by their nature tend to describe people and their actions. What is interesting is the relative importance of *she* and *he* (keyword ranks three and eleven respectively, with *he* following both *she* and *her*).

The use of pronouns is to some extent a feature of literary style. The literary subcorpus of the BNC shows a notably higher frequency than LOB's mixed prose, supporting the view that highly frequent personal pronouns are a stylistic feature. While in LOB's mixed prose, *he* is more than twice as frequent as *she*, the difference is much smaller in the BNC subcorpus. However, it must be borne in mind that the latter dates from thirty years later, and there have been some reductions in sexist terminology (Baker, 2008) and a slight increase in female and decrease in male pronouns (Baker, 2010) over the last few decades, even if these changes by no means indicate equality.

The relative frequencies of these two pronouns are interesting. (This study ignores *her* and *him*, because the former is both accusative and possessive, one word covering the meanings of two, *him* and *his*, making direct comparison difficult.) It is true that *he* is considerably more frequent than *she* also in *Adventure*, with *you* and *I* coming in between, but *she* is ranked fourteen as compared with thirty in LOB. It seems that female characters,

although still second to male, are more prominent in the adventure texts than in general prose of the time, challenging the popular view that girls in adventure fiction tend to be sidelined, made to stay at home while the boys enjoy the action.

Collocations

A glance at the relative frequencies of the words *girl, girls* etc. in the two corpora shows that it is only to be expected that the number of collocates will be very different (table 5). What is explored below is the collocations associated with all four lemmas GIRL, BOY, WOMAN, MAN in *Adventure*, but only with GIRL in *Girls*, as the very small numbers associated with the other search terms in *Girls* do not allow for much generalisation.

Table 5. Relative frequencies and number of collocates with the search terms

	girl	girls	boy	boys	woman	women	man	men
relative frequency in *Adv*	4.145	4.267	4.697	5.378	2.442	0.641	12.597	7.211
relative frequency in *Girls*	13.012	19.404	0.812	1.124	1.478	0.187	2.061	0.812
total collocates in *Adv*	39	35	36	40	21	4	100	73
total collocates in *Girls*	92	118	1	4	7	0	10	2

Although the ratio of female to male protagonists is often equal in the *Adventure* texts, the villains and minor characters are preponderantly male and adult. In *Girls* there is relatively little focus on adult characters, although since most of these are set in girls' schools one would expect most of these to be mistresses; possibly the slight majority of *man / men* in this genre is explained by the fact that the mistresses are generally referred to by name rather than as 'woman'.

Once function words such as pronouns, prepositions, determiners and conjunctions had been removed, certain categories could be observed, although not all the collocates fell into a clear category. As these were similar for both words (*girl, girls*) and lemmas (GIRL), the tables below outline the categories for the lemmas. Caldas-Coulthard and Moon (2010), following van Leeuwen (1996), identified adjectives used for identification (including 'classification' and 'physical') and appraisal. The common categories in my corpora were: age (=classification); appearance (= physical); appraisal; verbs (the last being a feature of the narrative genre). I have identified a further category, 'type', which includes both ethnicity and occupation (van Leeuwen's 'classification' and 'functionalisation' respectively). Whereas Caldas-Coulthard and Moon assigned words like lovely, beautiful to 'appraisal', I have put them in 'appearance' because they are often an intrinsic part of the description of female appearance. Words falling into a group I have termed 'adventure' collocate largely with MAN (and for the word *man* as opposed to the lemma this category also included *wounded, suddenly* and *police*). Table 6 below shows these categories (numbers are the rank of the collocate with that search term). Verbs and miscellaneous collocates are not included here as this would take us beyond the scope of this paper; verbs will be explored at a later date.

Table 6. Categories of collocates for GIRL/ BOY/ WOMAN/ MAN in *Adventure*
(Words in bold are shared between at least 2 search terms.)

Age		Appearance		Type		Appraisal		Adventure
GIRL								
little	5	pretty	40	french	8	**good**	43	**surprise**
elder	19	**tall**	47	english	14	silly		
young	21	lovely			15	**nice**		
younger	54	slim			16	**poor**		
older	61	beautiful			18	sensible		
					22	wonderful		
					34	ordinary		
					52	**naughty**		
					56	**brave**		
BOY								
small	41	**tall**	22	cornish	9	**naughty**	44	**surprise**
younger			72	**chinese**	38	**good**	66	rescue
little					45	**poor**		
elder								
WOMAN								
young	2	**dark**	4	**chinese**	7	evil		
old	3	plump	13	business	11	awful		
	5	**fat**						
	10	**tall**						
	16	**little**						
MAN								
old	9	**tall**	7	black	16	**good**	48	murdered
young	20	big			51	great	65	dead
elderly	24	**little**			77	**poor**	91	**surprise**
middle-	36	thin			92	**nice**	106	sick
aged	37	**dark**			96	bad		
	61	**fat**						
	85	fair						
	103	handsome						
	116	short						

Age-related terms are quite strong collocates of all the words, but there is a distinction between the children and the adults; *little* and *small* (and usually also *big*) when used of a child refer to age, but of an adult, to appearance. *Old* was also a collocate of the singular *girl* and *boy* (rank thirty-five and five respectively), but this is a feature of the period: 'old girl' or 'old boy' was a familiar form of address to a friend or sibling.

Other studies (e.g. Sigley and Holmes, 2002; Pearce 2008) have generally found more focus on physical appearance for females than males, and while this is true for GIRL vs BOY, the opposite is the case for WOMAN vs MAN. (The contrast is even stronger if only *man* rather than MAN is considered, with the addition of *swarthy, stout* and *well-dressed*.)

This is probably because, as mentioned above, most of the villains and minor 'good' characters tend to be male in this genre. It is certainly true, however, that there seems to be an emphasais on girls (though not women) being pretty, which is in line with others' findings.

It is obviously vital in this genre whether someone is a 'goodie' or a 'baddie', so it is not surprising that this comes up in the appraisal category; however MAN, unlike WOMAN, also seems to be appraised for other traits, most of them positive. Again, if the word *man* as opposed to the lemma is considered, this contrast is greater, with only three terms of appraisal collocating with *woman* but nine with *man*, including *brave, clever* and *kind.* The reverse is true for GIRL and BOY (and note that

here it is not a case of evil but of behaviour: *naughty* vs *good)*. Boys are noted for their behaviour or are described with compassion (*poor* is almost invariably an expression of pity rather than of lack of wealth), while girls attract far more appraisal, mostly positive.

The representation of girls in *Girls* was then compared with that in *Adventure*. Again (Table 7) only words are included which are clearly categorisable, with the exception of *old* and *my* (see below).

Table 7. Collocates of GIRL in *Girls*

Age		Appearance		Appraisal		Type		School-related	
6	elder	7	little	20	good	10	french	1	new
8	younger	[23	*big* – only	25	naughty	12	english	29	head
23	*big*	35	one	31	silly			102	school
120	older	54	example]	55	poor			143	matron
		66	tall	68	nice				
		104	pretty	69	kind				
		116	dark	74	bad				
			small	80	popular				
			slim	84	*strange*				

Big is generally used to describe the older girls in a school. *Strange* may be appraisal or simply mean 'unfamiliar'. *Poor* is always an expression of compassion. *Old* is sometimes, as in *Adventure*, used as a form of address (see 1), but 30% of instances are in the meaning of an alumna (2):

1) *Isabel laughed. 'Try again, old girl!' she cried. (NB to her twin sister)*
2) *Beside her sat Miss O'Ryan, an Old Girl of the school and now history mistress.*

Sigley and Holmes (2002) note the former use, and remark that, while it can be patronising, it may also indicate, as in (1), familiarity and / or solidarity. As *old* only collocates with *girl* and not with GIRL, it is not included in table 7.

The collocates in *Girls* correlate quite closely with those in *Adventure*. We see not just the same categories, but many of the same words. In both corpora, girls' appearance seems to be important, and this includes being pretty and slim. Also in both there are a number of collocates of appraisal, although a smaller proportion in *Girls* are positive (or a larger proportion negative). Auchmuty (1992) observed that girls' books offer the potential for a wider range of personalities than are sometimes found among female characters in other genres, and this finding may support that thesis. The school-related collocates, as with the school-themed keywords, are an

unsurprising feature and are balanced by the 'adventure' category of collocates in *Adventure*.

Three collocates of words rather than lemmas were explored through concordance lines: *my, only* and *you*. *My* collocates quite strongly with *girl* (rank 28, log dice 7.743) and is frequently a kind of patronising address, though not always male to female or older to younger person:

3) *'Get that, my girl?' 'Get what?' (a husband speaking to his wife)*
4) *'My dear girl! You're chums with Dorinda yourself now' (one schoolgirl to another)*

It may be that this is, like 'old girl / boy', a feature of dialogue of the period. The pattern is also found in *Adventure,* but (a) it is not exclusive to *girl*, being found also with *boy* (rank 8, log dice 8.079):

5) *'Look here, my boy...' (an adult to a child)*
6) *'You're suffering from post-operational depression, my boy.' (one man to a colleague)*

and (b) some of the examples in *Adventure* are from the books with teenage protagonists and refer to girlfriends:

7) *[David] thought, 'She's my girl. Ever since I first saw her...'*

Two other collocates deserve mention, namely *only* and *you*. *Only* collocates with *girl* in *Adventure* (rank thirty, log dice 6.557) and

50% of the examples are on the lines of 'she's only a girl':

8) *'...after all, you are only a girl, and ...'*

Only also collocates with *boy*, but never in this sense.

In general, pronouns were removed from the lists of collocates, but the concordance lines for *you+girls* in *Adventure* were examined. There is a popular conception of the boys telling the girls 'You girls stay here' whenever anything exciting is about to happen. Of forty-five occurrences, twenty-nine were the chunk *you girls* (and six were *you two girls*), often followed by an instruction, though there were only two examples of *you girls stay here*; other instructions included *you girls set the table; Go away, you girls*. However, there were also suggestions (9), questions (10) and approval / admiration (11).

9) *'I thought you two girls might stay here on guard...'*
10) *'...do you girls think you can find your way?'*
11) *'I must say you girls have got some pluck'*

It cannot be concluded that the female protagonists are necessarily being sidelined by being addressed as 'you girls', but it is certainly true that *you* did not collocate with *boys* in the same way, suggesting it is more frequent for boys to address girls, whether with instructions, questions or comments, than the reverse.

The collocations identified in this study do not give clear evidence of gender stereotypical discourse prosodies, with the obvious exception of the emphasis on physical beauty for girls, present in both corpora. This emphasis is in line with findings from numerous other studies (e.g. Pearce, 2008), but men's physical appearance is also of importance in *Adventure,* though not that of boys, or indeed of most adult male protagonists; it is the villains and assorted minor characters, largely male, who are frequently described. The far greater numbers of female characters in *Girls* allows for a wider range of personalities, some positive and some negative. 'Appraisal' collocates of girls in *Adventure* are largely positive; however there are considerably more of them than of boys. This may suggest that girls are somehow a

marked category in the genre, with boys being the unmarked norm.

Conclusion

Mid twentieth-century children's literature, as represented by these corpora, was not free of gender bias or gender stereotyping, not least in the fact that the different genres were explicitly targeted at 'boys and girls' or 'girls', respectively. However, there are actually considerable similarities in the keywords between the two genres studied here, and between the collocates of female and male characters in both.

The books in the *Adventure* corpus are aimed at a mixed readership and mostly feature both male and female protagonists, typically a group of girls and boys, often equal in number. However, this genre developed from adventure yarns aimed at male readers in the nineteenth and early twentieth centuries. In those earlier stories, female roles were generally confined to mothers and sweethearts, perhaps saying farewell to the hero in chapter one. It would therefore not be surprising if the female characters in *Adventure* had a largely passive role, kept out of most of the action, and indeed this is often the case with Enid Blyton, probably the best-known author in this selection. The fact that *he* is significantly more frequent than *she* in *Adventure* is a little disappointing; the fact that *she* is a keyword when *Adventure* is compared with LOB is in line with previous findings, and suggests that, while these authors did not present a world devoid of gender stereotypes, they did nonetheless give more prominence to female roles than they are sometimes credited as doing. There are discourse prosodies of physical beauty associated with female, and adventurous spirit with male characters, yet they also have also many collocates in common.

It is a clear limitation that these corpora are very small. The process of hand-scanning and digitising the books was laborious and time-consuming, limiting the size. However, the selected texts were representative of the much larger body used for content analysis.

They include works by the most popular and prolific authors of the period in the two genres. It is likely that these authors, if not these specific works, were read by any children of the time who were reading adventure and girls' books.

It is important to know how popular literature might have influenced child readers. From 1930 onwards girls' interest in action and adventure was being acknowledged with the publication of 'mixed' adventure books. While this study indicates some gender stereotyping (in reporting verbs, in the emphasis on being pretty, more male characters overall in *Adventure*), it does not support the view that there was clear, unrelenting polarisation of gender roles, and this is in line with the previous, more qualitative studies.

References

Anderson, D. A., Hamilton, M.C., Gilkson, G.M. and Palmer, S.K. (2023) Gender Socialization, Family Leadership and the Invisible Father in Children's Literature. *Journal of Family Issues 44 (5)*, pp.1343-1357

Auchmuty, R. (1992). *A World of Girls*. London: The Women's Press.

Baker, P. (2004). Querying Keywords: Questions of Difference, Frequency and Sense in Keywords Analysis. *Journal of English Linguistics 32(4),* pp. 346-359. http://dx.doi.org/10.1177/0075424204269894

Baker, P. (2008). *Sexed texts: Language, gender and sexuality.* London: Equinox.

Baker, P. (2010). Will Ms Ever Be as Frequent as Mr? A Corpus-based Comparison of Gendered Terms across Four Diachronic Corpora of British English. *Gender and Language 4 (1)*, pp. 125-149

Baker, P. (2014). *Using Corpora to Analyze Gender*. London: Bloomsbury.

Brezina, V. (2018). *Statistics in corpus linguistics*. Cambridge University Press

Brezina, V., Weill-Tessier, P. & McEnery, A. (2020). #LancsBox v. 6.0 [software]. Retrieved from http://corpora.lancs.ac.uk/lancsbox

Butts, D. (2018). Shaping boyhood: British Empire builders and adventurers. In Hunt, P. (Ed) *International Companion Encyclopedia of Children's Literature 2nd Ed Vol 1, (*pp. 340-351). London, New York: Routledge.

Butts, D. (2021) Biggles - Hero of the Air. In D. Jones and T. Watkins (Eds) *A Necessary Fantasy? The Heroic Figure in Children's Popular Culture.* (pp. 137-152). Lond, New York: Routledge.

Caldas-Coulthard, C. R. and Moon, R. (2010). 'Curvy, Hunky, Kinky': Using Corpora as Tools for Critical Analysis. *Discourse and Society 21(2):* pp. 99-133

Čermáková, A. and Mahlberg, M. (2021). Gender Inequality and Female Body Language in Children's Literature. *Digital Scholarship in the Humanities 36(2),* pp. 72-77 https://doi.org/10.1023/b:sers.0000018892.26527.ea

Eyre, F. (1972). *British children's books in the twentieth century*. London: Longman.

Gehrau, V., Brüggemann, T., & Handrup, J. (2016). Media and occupational aspirations: The effect of television on career aspirations of adolescents. *Journal of Broadcasting & Electronic Media, 60(3)*, pp. 465-483. https://doi.org/10.1080/08838151.2016.1203319

Gablasova, D., Brezina, V. & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing and Interpreting the Evidence. *Language Learning 67*(S1), pp. 155-179

Hunt, S. (2011). *The Discoursal Construction of Female Physical Identity in Selected Works in Children's Literature.* (PhD Thesis). Rhodes University

Hunt, S. (2015). Representations of gender and agency in the Harry Potter Series. In P. Baker and T. McEnery (Eds) *Corpora and discourse studies,* (pp. 266-284). Basingstoke: Palgrave Macmillan https://doi.org/10.1057/9781137431738.0017

Hunt, S. (2017). *Boast and Bellow, Giggle or Chatter: Gender and Verbs of Speech in Children's Fiction*. Retrieved from https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2017/general/paper306.pdf

Hunt, S. (2019). Sugar and Spice or Snails and Tails: The Gendered Representation of Emotion in Children's Fiction. (*International Corpus Linguistics Conference*) Cardiff University, Cardiff 22-26 July 2019.

Koller, V. and Mautner, G. (2004). Computer applications in critical discourse analysis. In C. Coffin, A. Hewings and K. O'Halloran (Eds) *Applying English grammar: Functional and corpus approaches*, (pp. 216-228). Abingdon: Routledge.

Nodelmann, P. (2008). *The hidden adult: Defining children's literature*. Baltimore, Maryland: Johns Hopkins University Press.

Pearce, M. (2008). Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine. *Corpora 3*(1), pp. 1-29 http://dx.doi.org/10.3366/E174950320800004X

Poynter, E. (2018) *You girls stay here: Gender roles in popular British children's adventure fiction, 1930-70*. Newcastle: Cambridge Scholars Publishing

Poynter, E. (2020) Aggressive but loyal: modification and gender roles in British children's adventure fiction. *Gender and Language 14 (2*), pp. 175-196

Rayson, P. (2008). From Key Words to Key Semantic Domains. *International Journal of Corpus Linguistics 13(4),* pp. 519-49. http://dx.doi.org/10.1075/ijcl.13.4.06ray

Rowe Townsend, J. (1990). *Written for children*. London: Bodley Head.

Ruano, P. (2018). An analysis of Charles Dickens's Gender-based Use of Speech Verbs. *Gender & Language 12(2),* pp. 192-217. http://dx.doi.org/10.1558/genl.31137

Rudd, D. (2000). *Enid Blyton and the mystery of children's literature*. Basingstoke: Palgrave.

Rudd, D. (2009). In Defence of the Indefensible? Some Grounds for Enid Blyton's Appeal. In J. Maybin and N. J. Watson (Eds) *Children's Literature: Approaches and Territories.,* (pp. 168-182)*.* Milton Keynes: Palgrave Macmillan.

Sigley, R. and Holmes, J. (2002). Looking at *girls* in Corpora of English. *Journal of English Linguistics 30(2)*, pp. 138-157

Skočajić, M. M., Radosavljević, J. G., Okičić, M. G., Janković, I. O., & Žeželj, I. L. (2020). Boys just don't! Gender stereotyping and sanctioning of counter-stereotypical behavior in pre- schoolers. *Sex Roles, 82* (3–4), pp. 163–172. https://doi.org/10.1007/ s11199-019-01051-x

Sunderland, J. (2011). *Language, gender and children's fiction.* London: Continuum International Publishing Group.

Tsao, Y-L. (2020) Gender issues in young children's literature. *Reading Improvement 57* (1), pp.16-21

Van Leeuwen, T. (1996). The Representation of Social Actors, in C.R. Caldas-Coulthard and M. Coulthard (Eds) *Texts and Practices: Readings in Critical Discourse Analysis*,(pp. 32–70). London: Routledge.

Zanfabro, G. (2017). Gender Matters: What Is at Stake in Dealing with Children's Literature? *TRANS* Retrieved from http://transrevuews.org/1440

APPENDIX: The corpora

'*Adventure*'

Author	Title	Date of 1st publication	No. of words
M. E. ATKINSON	Smugglers' Gap	1939	72,023
	The Monster of Widgeon Weir	1943	91,750
ENID BLYTON	The Island of Adventure	1944	64,040
	The Mountain of Adventure	1949	59,335
ELINOR M. BRENT-DYER	Chudleigh Hold	1954	69,440
	Top Secret	1955	55,311
GWENDOLINE COURTNEY	The Grenville Garrison	1940	71,123
	Mermaid House	1953	55,095
CAPT. W. E. JOHNS	Worrals Carries On	1942	37,896
	Worrals in the Wastelands	1949	51,899
	Biggles 'Fails To Return'	1943	57,498
	Biggles Looks Back	1965	50,177
ARTHUR RANSOME	Swallowdale	1931	121,172
	Missee Lee	1941	92,677
MALCOLM SAVILLE	The Secret of Grey Walls	1947	87,107
	Rye Royal	1969	63,229
	The Purple Valley	1964	72,632
	Power of Three	1968	69,456
Total number of words			**1,241,860**

For the digital corpus two books were selected from each author, or in the case of Johns and Saville, four, two from each of their adventure genres (male or female protagonists and children's or Young Adults', respectively).

'*Girls*'

Author	Title	Date of 1st publication	No. of words
ENID BLYTON	The O'Sullivan Twins	1942	44,153
	Claudine at St Clare's	1944	45.384
ELINOR M. BRENT-DYER	Jo Returns to the Chalet School	1936	66,830
	The Chalet School Does It Again	1955	69,070
GWENDOLINE COURTNEY	Sally's Family	1946	77,642
	At School with the Stanhopes	1951	58,033
DORITA FAIRLIE-BRUCE	The School on the Moor	1931	65,676
	The School in the Woods	1940	57,521
Total number of words			**484,309**

Three of the women who wrote adventure fiction also wrote for girls; to increase the size and scope of this corpus two works by another well-known writer for girls of the period were added. Some possible authors / books for girls had to be rejected because they contained no male characters, which was not desirable as no comparison between the genders would have been possible.

For both corpora, the books selected for each author were where possible an early one from a series and then a later one from the same series.