



A Diachronic Study of Gendered Body Language in French-Language Children's Literature from the Nineteenth and Early Twentieth Centuries

David Korfhagen and Theresa McGarry
East Tennessee State University

INTRODUCTION

The speech, behavior, and body positioning/configuration of fictional characters communicate societal norms and ideologies, most often implicitly. Accordingly, analysis of the language involving body and voice in literature can inform our understanding of the norms and beliefs to which readers are exposed. Focusing on gender ideology and child readers, Čermáková and Mahlberg showed that repeated word clusters containing body part lexical items with gendered possessive pronouns, such as “her arms” or “his face,” evidenced a gender binary in British children's literature of the nineteenth century that became less distinct in that of the twentieth century. Attempting to extrapolate from these findings to French literature, this study examines verbal patterns involving body part and voice terminology in French children's books of the nineteenth and twentieth centuries. The concept of gender we assume is the performative view, in which gender is fluid and dynamic, performed by repeated acts, gestures, and desires (Butler). We demonstrate that, in the nineteenth century, the recurring verbal clusters involving body and voice terms manifest the ideological gender binary. In the twentieth century, though some changes appear, contrary to the findings of Čermáková and Mahlberg the overall picture is more one of continuation, at least through the first half of the century.

SEXISM IN CHILDREN'S LITERATURE

As a genre, books written for children exhibit sexist tendencies. A prominent way in which they do so is by the preponderance of male characters, particularly in key roles. To cite just two of many studies, Hamilton et al. found that, in 200 award-winning and/or top-selling picture books in the US from around the beginning of the twenty-first century, title and main characters were male nearly twice as often as they were female. Similarly, McCabe et al. showed that, in twentieth-century children's books published in the United States, male characters occurred at nearly twice the rate of female characters, and central characters were male 1.6 times as often as they were female. In addition to outnumbering female characters, male characters engage in different activities. Occupations are stereotyped (Hamilton et al.), and many studies

such as Pownall and Heflick observe a tendency for female characters to engage in domestic activities while male characters pursue adventure. Adams and Harper cite numerous studies showing that female characters, in addition to being underrepresented, are “more likely to be restricted to domestic and nurturing social roles and [to exhibit] disproportionately low levels of authority, competence and social status” (456).

The aggregate story world created in children's books has been characterized as “a microcosm of ideologies, values, and beliefs from the dominant culture, including gender ideologies and scripts” to which we expose children (Taylor 301). The effects of their contact with this microcosm have been repeatedly demonstrated. Filipović concludes from a literature review that gender stereotyping affects child readers' “activity choices, career aspirations, and academic outcomes” and negatively affects boys as well as girls by “limiting the scope of acceptable roles and behaviors, and normalizing gender inequalities” (311). Similarly, Adam and Harper summarize previous research as showing that literary depictions of gender can affect children's career aspirations, attitudes about future societal roles, and personalities, and gender bias in books strengthens children's biases, raises boys' entitlement while lowering girls' self-esteem, and teaches that girls are less valuable.

What makes this kind of transmission particularly effective is the way that children tend to learn things like language and ideology, which involves a significant degree of probabilistic learning. Without taking a position on whether or not child learners are born with inherent language structures, it is safe to assert that they are sensitive to patterns in the data they encounter. Chater & Manning argue this point clearly with regard to language learning, and it is easy to see that it extends to other areas. For example, children in the U.S. generally learn without explicit instruction that some adults drink coffee regularly but children do not, though they may see some exceptions. A great many gendered norms, relating to dress, toys, hobbies and interests, playgroup membership, speech patterns, etc., are learned implicitly in this way, by observation of patterns leading to induction. The attributes and behavior of gendered characters in the books children read, especially the realistic ones, also constitute input to their induction processes. Therefore, it is important to examine broad samples of such literature to ascertain what kind of world it constructs across the books children read, by means of the patterns to which it exposes them.

For this task, a corpus methodology has obvious advantages and has already been demonstrated to be informative. The example that inspired the current study is Čermáková & Mahlberg's investigation of British children's literature using two large corpora: the 4.4 million-word ChiLit, containing nineteenth-century works, and a 12.9 million-word subcorpus of the Oxford Children's Corpus, OCC2000+, which contains twentieth-century books. Starting from the premise that body language goes beyond describing characters' bodies to showing how they interact with each other and with their material surroundings, the researchers set out to identify gendered differences in the mention of body parts within each century and then compare the differences between the two centuries. Following the methodology established in Mahlberg, the comparison focused on recurring five-word clusters that included a noun referencing a body part following the pronoun *her* or *his*, for example *his hands in his pockets* or *her face with her hands*.

Their results indicated that in the nineteenth century the patterns of the clusters corresponded to a gender binary of femininity and masculinity. The differences were mostly of degree, but a few clusters did categorically associate with one pronoun. The vast majority of those that did associate with *his*, which is to say they were associated with male characters. The clusters including *her* were fewer and less varied. Further, they tended to show the female characters in relational bodily actions and configurations, such as throwing their arms around

someone else, while the clusters with *his* tended to show the character in an independent action or posture. For example, the single most common cluster with *his* was *his hands in his pockets*, and the counterpart *her hands in her pockets* did not occur at all. The language of touch was prominent and indicated both proximity of characters and asymmetry, such as the touch of a hand on another person's shoulder. Finally, in clusters with the word *eyes*, tears were used to indicate emotions for both the genders portrayed, but over the course of the century they became less frequent for male characters, which paralleled social changes in the ideology of maleness.

In the twentieth century, the researchers found, the distinction lessened. The main mechanism of the convergence was an increase in the variety of clusters containing *her*, including a smaller number of clusters exclusive to males. It appeared that there were more options for female characters to enact femininity, including the emergence of a new and prominent cluster, *her hands on her hips*, which is independent rather than relational and can be interpreted as a posture of dominance. A change also occurred in the language of touch, in that the texts included more thoughtfulness on the part of both the agent and the recipient of the touch.

In general, Čermáková & Mahlberg noted that the five-word clusters, hereafter referred to as five-grams, were culture-dependent, as is body language in general. This rather unsurprising observation inspires the question of to what degree the findings can extrapolate to other cultures. French literature, like British, includes a substantial children's literature component from both the nineteenth and the twentieth centuries and therefore constitutes a good subject for exploratory extrapolation of the findings.

RESEARCH QUESTIONS

Čermáková & Mahlberg's methodology depends on the English locution in which body parts are commonly referred to with a genitive pronoun in determiner function. The gender of the body part possessor is indicated by the genitive pronoun *her* or *his*, and the phrases can thus be easily sorted. The French situation is different in that the gender of the genitive pronoun in determiner position is grammatical – not social – gender, matching the grammatical gender of the body part rather than of its possessor, or, in the case of plurals, not indicating gender. For example, *son épaule* is used to refer to the shoulder of either a female or a male, and *ses épaules* translates to either 'her shoulders' or 'his shoulders.' Further, it is common in French to refer to body parts with the definite article rather than the genitive pronoun determiner. For example, to express 'he had his book under his arm' it would be more common to say *il avait son livre sous le bras*, literally 'he had his book under the arm' than to say *il avait son livre sous son bras* 'he had his book under his arm.' For these reasons, it is not possible to sort occurrences of gendered body parts by genitive pronoun determiner. Thus, our first research question is as follows:

1. What methodological adaptations are necessary to study body part clusters in French?

The next question follows more straightforwardly from the goal of extrapolating from Čermáková & Mahlberg's findings:

2. Does body part language in nineteenth- and twentieth-century French children's literature pattern like British? What other tendencies might it show?

Our third question follows from a preliminary examination of the five-grams appearing in the data. The voice is not, strictly speaking, a part of the body, but it does seem to be in some sense a physical attribute. The term appeared with sufficient frequency that ... Recurring clusters suggested that characterizations of voice might also be important in the construction of gender, so, while voice terms were not addressed in Čermáková & Mahlberg's analysis, we also posed the following question:

3. Does voice terminology also contribute to gendered construction of characters?

METHODOLOGY

3.1 DATA

Despite the substantial body of French children's literature created in the nineteenth and twentieth centuries, large corpora parallel to those existing for English have yet to be built. To construct a searchable corpus for our analysis, we took advantage of electronic texts freely available online, mainly at gutenberg.org but also including some from ebooksgratuit.com. We gathered only prose narratives of books originally published in the nineteenth and twentieth centuries, and only books in the original French, to avoid possible translation issues. In order to escape undue influence from the style of prolific authors, we took only one book from any given author. The main disadvantage to restricting the corpus in this way is that unequal influences correlating to differences in popularity among authors is not reflected. However, in a relatively small corpus such as ours, it is important to guard against results reflecting possible authorial idiosyncrasies rather than trends. While we acknowledge that certain authors, such as Jules Verne, might have had more influence than others, our intention was to gauge the broad societal trends, and we felt that, given how small our corpus was, including multiple works by authors. Future research will hopefully involve corpora, like those used by Čermáková and Mahlberg, that are big enough to accommodate this kind of consideration.

Because of copyright issues, it turned out to be unfeasible to obtain books from after 1945, so our twentieth-century corpus in fact represents only the first half of that century, and it is about half the size of the nineteenth-century sub-corpus, a limitation of the study that we hope to address in the future. Table 1 shows the sizes of the subcorpora. While they are clearly much smaller than the corpora used by Čermáková & Mahlberg, we believe that they are large enough to yield meaningful results.

	Books	Approximate words
19 th century	46	1,800,500
1901-1945	19	900,000

Table 1. Corpora of French children's books used for analysis

3.2 IDENTIFICATION OF FIVE-GRAMS

We identified all five-grams occurring in the text by means of the n-grams tool in AntConc, a free download that allows separate files, such as e-books, to be combined into corpora and generates n-grams of any desired size and range. Having generated a list of examples for each century, we manually sorted out those contained a body-part term or a term relating to voice. The software provides a certain amount of contextualization that allows for such sorting, and for identifying the gender of the character, as shown in Figure 1.

Left context	Hit	Right context
<i>répond Fanni en</i>	<i>se jetant à son cou</i>	<i>si vous étiez assez bonne</i>

Figure 1. Example of five-gram as shown in AntConc

The five-gram appears in the middle column, labeled Hit. The column labels Left and Right context show the language occurring in the text before and after the five-gram. The number of words to be included in these columns is adjustable, and when the context shown is insufficient, more can easily be seen by choosing a different view from the menu.

We discarded instances in which the body part did not belong to a gendered human character, for example, those in which it belonged to an animal. Having much smaller corpora than Čermáková & Mahlberg required adjusting the minimum thresholds of occurrence for any particular five-gram to be considered. For the nineteenth century, we set the minimum at seven and obtained 36 distinct recurring clusters. For the twentieth century, because the subcorpus was half the size of the nineteenth-century one, we reduced the minimum to four, which yielded 17 distinct clusters. Doing this allowed approximately the same range of cluster types to be considered across the two corpora. In each case, we set the minimum range at two, which is to say the cluster had to occur in more than one book, to obviate the possibility of dealing with an idiosyncrasy of any one author.

We added to Čermáková & Mahlberg's methodology a norming step meant to partially control for the issue that male characters predominate in children's books; to alleviate the possibility that a difference in the frequency of any five-gram may just be fallout from that imbalance in character gender, we calculated the ratio of the occurrences of *il* 'he' compared to *elle* 'she' in each subcorpus, filtering out instances in which *il* was used in reference to inanimate objects or in impersonal constructions, such as *il y a* 'there is/are'. The expected predominance of the male pronoun was pronounced: the ratio was more than 1.62:1 in the nineteenth century, and the discrepancy increased in the twentieth to 2:1. This proportion was then available to be used as a baseline against which to measure discrepancies in the occurrence of any five-gram. Examples are shown in the Results section, below. Table 2 shows the number of distinct five-grams obtained, the total number of occurrences of those five-grams aggregated, and the baseline ratio of *il* 'he' to *elle* 'she' for each subcorpus.

	19 th century (freq. ≥7, range ≥2)	20 th century (freq. ≥4, range ≥2)
five-gram types	36	17
five-gram hits	364	81
baseline ratio (<i>il</i> ‘he’: <i>elle</i> ‘she’)	1.62:1	2.01:1

Table 2. Five-grams above the thresholds and ratios of *il* ‘he’ to *elle* ‘she’

RESULTS

4.1 METHODOLOGICAL ADAPTATIONS

The first research question concerned the feasibility of extending Čermáková & Mahlberg’s methodology to a language in which direct linguistic parallels of the phrases “her/his [body part]” do not exist. For the English corpora, the researchers were able to generate the list of five-grams occurring above the thresholds and then extract the ones that contained either “her” or “his” and a body part noun, and the five-gram was automatically gendered. In French, as discussed above, a genitive pronoun in a phrase such as *son cou* expresses grammatical rather than social gender. Further, the frequency of preceding the body part with a definite article rather than a genitive pronoun required including phrases with that structure also, as to ignore them would have been to omit a significant component of gendered references to body parts.

The adapted procedure involved a few more steps, as we expected, but was not overly cumbersome. We generated lists of five-grams over the thresholds described above, in the methodology section, setting the left context size to 20 words and the right context size to 15 words. When we had manually extracted all those containing a body part or voice term, including with each item the left and right context, the text was in most cases sufficient to identify the relevant gender for each body part referred to. Figure 2 shows the example of an occurrence of *bras autour de son cou* ‘arms around her/his neck.’ The left context is sufficient to determine that the arms belong to a character gendered female, and the neck belongs to a character gendered male.

Left context	Hit	Right context
<i>mains. Tous trois pleuraient. -- Mon père, dit Christine en se mettant à genoux près de lui et en passant un</i>	<i>bras autour de son cou,</i>	<i>pendant que de l'autre main elle tenait celle de François, mon père, votre chagrin,</i>

Figure 2. Example of a hit with 20-word left context and 15-word right context

In other cases, the context was insufficient. However, our corpus was under no restrictions limiting the amount of text we could view; the entirety of each e-book was available, and any hit selected could be immediately located in its document with the AntConc function File View.

Thus, we were able to classify five-grams as male-gendered, female-gendered, or both, and to discard nonhuman ones, relatively easily and quickly.

In addition, the procedure we used yielded a more inclusive sample than that of Čermáková & Mahlberg in several ways. For one, we were able to include body part terms where the possessor was first person or third person plural. Moreover, when the reference was plural, we were in most cases able to count the possessors precisely. For example, *des larmes dans les yeux* '[with] tears in their eyes' occurred nine times in the nineteenth-century data. In two occurrences, the eyes of more than one character were at issue, and we were able to determine that in one of those occurrences both characters were female, and in the other one character was female and the other male.

Clearly, the larger the corpus, the more time and effort will be required for this classification step of the analysis. Čermáková & Mahlberg's nineteenth- and twentieth-century corpora total about 6.5 times as many words as ours, so the time required would have been significantly longer, but not, we believe, prohibitive. Further, we submit that our results demonstrate the meaningful analyses can be obtained with relatively small corpora.

4.2 PATTERNS IN GENDERED BODY PART FIVE-GRAMS

Our second research question concerned the occurrences of the five-grams in relation to gendered characters. Table 3 shows the five-grams above the thresholds for the nineteenth century. The second column indicates the total frequency, and the third shows the number of books in which the sequence occurs. The fourth and fifth columns show the gender of the characters to whom the locution refers, and the last column shows the ratio of male:female usages normalized to a denominator of 1. Note that the baseline ratio established by comparing instances of *il* to instances of *elle*, as explained in the methodology section above, is 1.62:1. Thus, normalized ratios above this baseline indicate five-grams skewed male, and those below this baseline indicate skewing towards female.

5-Gram	Freq.	Range	M	F	Ratio
<i>se jeta dans les bras</i>	18	10	13	5	2.6:1
<i>de la tête aux pieds</i>	13	10	9	6	1.5:1
<i>dans les bras de son/sa</i>	20	15	15	5	3 :1
<i>du coin de l'œil</i>	16	12	8	6	1.33:1
<i>sur la pointe des pieds</i>	14	8	9	6	1.5:1
<i>jeta dans les bras de</i>	13	8	8	5	1.6:1
<i>la tête dans ses mains*</i>	13	9	8	3	2.67:1
<i>dit il/elle d'une voix</i>	21	10	11	10	1.1:1
<i>les yeux pleins de larmes</i>	12	9	5	8	0.625:1
<i>jeta un coup d œil</i>	11	8	8	3	2.67:1
<i>et dit d'une voix</i>	10	9	9	1	9:1
<i>t il/elle d une voix</i>	14	11	10	4	2.5:1
<i>un coup d'œil sur</i>	10	9	7	2	3.5:1
<i>aime de tout mon cœur</i>	9	6	4	5	0.8:1
<i>dans ses bras et l'</i>	9	4	4	5	0.8:1
<i>des larmes dans les yeux</i>	9	6	5	5	1:1

Gendered Body Language in French Children's Literature

<i>fit un signe de tête</i>	9	7	6	3	2:1
<i>la main sur la tête</i>	9	6	8 main, 7 tête	1 main, 2 tête	8:1 main, 3.5:1 tête
<i>prit dans ses bras et</i>	9	9	5	4	1.25:1
<i>sa tête dans ses mains</i>	9	9	7	2	3.5:1
<i>se jetant à son cou</i>	9	7	4	5	0.8:1
<i>d un coup de pied</i>	8	5	6	2	3:1
<i>des pieds à la tête</i>	8	6	5	3	1.67:1
<i>dit il d'un ton</i>	8	7	exclusively male		
<i>en se frottant les mains</i>	8	6	exclusively male		
<i>jeter un coup d'œil</i>	8	6	3	1	3:1
<i>les mains derrière le dos</i>	7	5	exclusively male ¹		
<i>prit par la main et</i>	8	6	6	2	3:1
<i>sur l'épaule de son/sa</i>	9	6	8	1	8:1
<i>bras autour de son cou</i>	7	6	0 bras, 5 cou	7 bras, 2 cou	bras exclusively female, cou 2:5:1
<i>dans ses bras et lui</i>	7	5	3	4	0.75:1
<i>de sa voix la plus</i>	7	6	4	3	1.33:1
<i>le/la prit par la main</i>	10	7	7	3	2.33:1
<i>les coudes sur la table</i>	7	6	exclusively male		
<i>se jeta dans ses bras</i>	7	5	5	1	5:1
<i>tête dans ses mains et</i>	7	7	5	2	2.5:1
<i>totals</i>	368	266	256	117	2.189:1

Table 3. Nineteenth century - baseline ratio: 1.62:1

First, we asked whether body part language in nineteenth- and twentieth-century French children's literature patterns similarly to the English terms in Čermáková & Mahlberg's data. It emerged that, for the nineteenth century, our results do largely parallel theirs. As with the English books, we found that the differences between the descriptions of male and female characters are primarily a matter of degree, not category. For example, the most common body-part cluster, *se jeta dans les bras* 'threw herself /himself into the arms' skews male. The ratio of 13:5 can be restated as 2.6:1. Since this ratio is higher than the baseline ratio of 1.62:1, the discrepancy is not explainable simply by the preponderance of male characters; rather, it suggests that characters throw themselves into the arms of male characters considerably more often than the arms of female characters. The second most common five-gram, *de la tête aux pieds* 'from head to foot,' skews female, at 9 male:6 female. Normalized, this ratio is 1.5:1, which is slightly less than the baseline ratio of 1.62:1, suggesting that the usage is more often associated with female gender. Most differences are preponderances of this type.

However, also in line with the English data, categorical usages do exist, and they mostly characterize males. Further, the male-associated ones tend toward describing independently achieved actions, while the female ones are generally more interpersonal, which also repeats a pattern in the British books. As shown in Table 3, above, of the 36 prevalent five-grams, four are

¹ In one instance of *les mains derrière le dos* out of the seven identified, the relevant characters were a group of children, with gender unspecified, presumably mixed. Discounting this instance, the usage is exclusively male.

exclusive to male characters. Three of them, *en se frottant les mains* ‘rubbing his hands,’ *les mains derrière le dos* ‘his hands behind his back,’ and *les coudes sur la table* ‘his elbows on the table,’ are independent postures, and the other, *dit-il d’un ton* ‘he said in a tone,’ concerns voice, discussed in Section 4.3, below, and is also independent of other characters. Only one five-gram is exclusively female: in the locution *bras autour de son cou*, ‘arms around her/his neck,’ all instances of arms refer to arms of females. It is the female characters whose arms are around the necks of other characters, and this configuration is, of course, interpersonal. Thus, the nineteenth-century French books appear to instantiate the gender binary in ways quite similar to those of the English books.

When comparing the nineteenth and twentieth centuries, however, our findings diverge from those of Čermáková & Mahlberg. Their study showed an increase in space and variety among descriptions of female characters during the transition from one century to the next, resulting in more similarity between genders. Ours, however, showed the reverse. Table 4 shows the trend in terms of clusters that were associated with only one of the genders; while our nineteenth-century data show the expected prevalence of male-only clusters over female-only ones (4:1), we find a slight increase in frequency of male-only clusters in the twentieth-century data and an absence of female-only ones.

	Male-only	Female-only
19 th	4; none relational <i>en se frottant les mains</i> ‘rubbing his hands’ <i>les coudes sur la table</i> ‘his elbows on the table’	1 <i>bras autour de son cou</i> ‘[her] arms around her/his neck’
20 th	7; none relational <i>au dessus de sa tête</i> ‘above his head’	N/A

Table 4. Gender-exclusive cluster frequency

Results shown in Tables 5 and 6 seem to further exemplify the trend. In expressions involving *bras* ‘arm(s),’ for the nineteenth century, we see a general tendency toward parity between the two genders, and in some cases a favorability toward female characters, given the baseline male:female ratio of 1.62:1. For example, in *se jeta dans les bras* ‘threw herself /himself into the arms,’ the ratio of 13 male body part clusters (MBPC) to 5 female body part clusters (FBPC) equals 2.6:1. Likewise, the ratio for all of the clusters combined is 58:33, or 1.76:1. Given the smaller number of examples for the 20th century, shown in Table 6, the 3:1 ratio is less significant, but what is notable is the drastic decrease in the variety of expressions related to *bras* and the absence of the only female-exclusive cluster that we found across the corpora, *bras autour du son cou*.

Cluster	Frequency	Range	MBPC	FBPC
<i>se jeta dans les bras</i> 'threw herself /himself into the arms'	18	10	13	5
<i>dans les bras de son</i> 'into the arms of her/his'	15	11	15	0
<i>jeta dans les bras de</i> 'threw into the arms of'	13	8	8	5
<i>dans les bras et l'</i> 'into the arms and'	9	4	4	5
<i>prit dans ses bras et</i> 'took her/him in(to) her/his arms and'	9	9	5	4
<i>bras autour de son cou</i> 'arm(s) around her/his neck'	7	6	0 arm(s), 5 neck	7 arm(s), 2 neck
<i>dans ses bras et lui</i> 'in(to) her/his arms and to her/him...'	7	5	3	4
<i>se jeta dans ses bras</i> 'threw herself /himself into her/his arms'	7	5	5	1
<i>totals</i>	85	58	58	33

 Table 5. Nineteenth-century frequencies of expressions involving *bras* 'arm(s)'

Cluster	Frequency	Range	MBPC	FBPC
<i>au bout de son bras</i>	4	3	3	1

 Table 6. Twentieth-century frequencies of expressions involving *bras* 'arm(s)'

Perhaps more clear-cut are clusters involving *larmes* 'tears.' The nineteenth-century data (Table 7) appear to show a clear prevalence of female five-grams, given the ratio of 10:13, which normalizes to 0.78:1; taking into account the baseline ratio of 1.62 male:1 female only increases the weight in favor of female five-grams. This finding is not surprising, given stereotypes of women and girls as being more emotionally fragile than their male counterparts, making them more likely to be associated with shedding tears. At a first look, the twentieth-century data (Table 8) could possibly suggest that the stereotype has become less pervasive, at least as reflected in the texts covered in our corpora; the male:female ratio jumps from 0.78:1 in the

nineteenth century to 4:1 in the twentieth (again, with the caveat that the smaller corpus size for the twentieth century renders any conclusions tentative at best). However, it is notable that this context – closely associated with female characters due to the stereotype about their emotional fragility – occurs less frequently than in the preceding century. Again, this finding contrasts with the increase in space and variety among descriptions of female characters found by Čermáková & Mahlberg.

Cluster	Frequency	Range	MBPC	FBPC
<i>les yeux pleins de larmes</i> 'her/his eyes full of tears'	12	9	5	8
<i>des larmes dans les yeux</i> 'tears in her/his eyes'	9	6	5	5

Table 7. Nineteenth-century frequencies of expressions involving *larmes* 'tears'

Cluster	Frequency	Range	MBPC	FBPC
<i>les yeux pleins de larmes</i> 'her/his eyes full of tears'	5	3	4	1

Table 8. Twentieth-century frequencies of expressions involving *larmes* 'tears'

Part two of our second research question asked what additional tendencies body part language in nineteenth- and twentieth-century French children's literature might show, beyond the tendencies Čermáková & Mahlberg highlight in their study. One of the more frequent five-grams in our data set was *de la tête aux pieds* 'from head to toe' which occurred eighteen times. *Des pieds à la tête* 'from toe to head' was also frequent, at thirteen occurrences. If considered two variants of the same locution, they would constitute the most-frequent cluster, the counterpart of which did not figure in the English analysis at all. We therefore examined their usage more closely in order to understand the way in which they contributed to character development. What emerged from this deeper analysis was a taxonomy of uses, categorized by association with a psychological/emotional condition, involving either a state or an identity, or with a physical condition or an event, involving either appearance or experience as the object of an action or a gaze. The examples in Table 9 illustrate these four possibilities.

Type	Example(s)
(1) State	<i>Deux des coupables rougirent soudain de la tête aux pieds.</i> 'Two of the culprits suddenly blushed from head to toe.'
(2) Identity	<i>Thérèse Brandi, une fière fille, une vraie Corse, et de la tête aux pieds.</i> 'Theresa Brandi, a proud daughter/girl, a true Corsican, from head to toe. ... Or ses bâtards, à lui, ne l'étaient pas qu'à demi, ils l'étaient des pieds à la tête et comme ceux du roi de France. 'However, his own were not just half bastards, they were bastards from head to toe and like those of the king of France.'

(3) Appearance	<i>Derrière le premier groupe d'amis, marcherait un guide, écarlate des pieds à la tête ...</i> ² 'Behind the first group of friends would walk a guide, scarlet from head to toe ...'
(4) Object	<i>Simplicie l'examina des pieds à la tête ...</i> 'Simplicie examined her from head to toe ...'

Table 9. Examples of psychological/emotional and physical conditions

The expression *de la tête aux pieds* in example (1) refers to the overwhelming nature of the characters' emotional state, in (2) it refers to the extent of the characters' identities, in (3) it refers to the character's physical presentation, and in (4) the character is viewed as an object of a visual examination.

Comments about female characters remained roughly on a parity with those about their male counterparts with respect to psychological/emotional characteristics into the twentieth century; overall, the male:female ratio was 6:4, or 1.5:1, compared to the baseline ratio of 2:1. In the twentieth century, these usages did not reference physical characteristics at all for either gender, as shown in Table 10. Moreover, our data suggest that physical characteristics more frequently associated with male characters than female ones, at an overall ratio of 10:4, or 2.5:1. In sum, the tendencies shown here contradict our expectation of more objectification of female characters than male, and they represent the only case in which our findings could be construed as agreeing with those of Čermáková & Mahlberg.

	19th			20th	
	MBPC	FBPC		MBPC	FBPC
State	2	5		4	2
Identity	0	0		2	2
Appearance	5	0		0	0
Object	5	4		0	0

Table 10. Frequencies of *de la tête aux pieds* / *des pieds à la tête* 'from head to toe / from toe to head'

At this point we can only speculate on the reasons for the differences between our findings and those of Čermáková & Mahlberg. An obvious direction would be to investigate societal trends advancing women's rights and position in society. A clear possibility would relate to differences in conceptions of acceptable femininity between French and British societies in the early twentieth century. It might be fruitful to look at the timing of societal phenomena such as women's suffrage. In the UK, all women had the right to vote by 1928, while in France this was not achieved until 1944. Future studies might address this issue by expanding the corpus to the second half of the twentieth century and by breaking the data down into smaller increments

² Note that *écarlate* 'scarlet' refers here to the color of the guide's clothing; thus, this is not a metonymic expression conveying the character's emotional state, along the lines of 'she/he was red in the face.' (Cf. the example for State in Table 9).

that might correlate with significant events that afforded women greater agency, such as taking jobs in factories during the World Wars, etc.

4.3 VOICE TERMINOLOGY

Our third research question was whether terminology related to *voix* ‘voice’ also contributes to gendered construction of characters in nineteenth- and twentieth-century French children’s literature. This question was inspired by the multiple high-frequency variants involving the word that appeared in the preliminary examination of the five-grams generated, shown in Tables 11 and 12. The locutions begin with or follow a verb of oral expression such as *dit* ‘said,’ *s’écria* ‘cried,’ or *répondit* ‘answered.’ In each case, the word *voix* is followed in the data by an adjective phrase, such as *tonnante* ‘thundering,’ *la plus douce* ‘sweetest,’ or *si forte que toutes les vitres de la salle se brisèrent en éclat* ‘so strong that all the window panes in the room shattered.’ Thus, the characterization involves a particular quality of voice.

Cluster	Frequency	Range	MBPC	FBPC	Ratio
<i>dit il/elle d'une voix</i> ‘he/she said in a voice’	21	10	11	10	1.1:1
<i>et dit d'une voix</i> ‘and [he/she] said in a voice’	10	9	9	1	9:1
<i>t il/elle d'une voix³</i> ‘he/she [said] in a voice’	14	11	10	4	2.5:1
<i>de sa voix la plus</i> ‘in his/her loudest voice’	7	6	4	3	1.33:1
<i>totals</i>	52	36	34	18	1.89:1

Table 11. Nineteenth-century frequencies of *voix* ‘voice’

Cluster	Frequency	Range	MBPC	FBPC	Ratio
<i>dit il/elle d'une voix</i> ‘he/she said in a voice’	6	4	male-exclusive		
<i>s’écrie d'une voix</i> ‘shouts in a voice’	6	1	male-exclusive		
<i>de sa voix la plus</i> ‘in his/her loudest voice’	4	3	2	2	1:1
<i>totals</i>	16	6	14	2	7:1

³ The lone *t* is a required phonological insertion between the 3rd-person singular simple-past verb form and *il* or *elle*. Thus, these occurrences represent cases where *il* or *elle* is the subject of a reportative verb, such as *cria*. AntConc registered this element as a lexical item when producing the 5-grams reported here.

Table 12. Twentieth-century frequencies of *voix* 'voice'

In the nineteenth century, individual five-gram types occur at ratios both below and well above the baseline of 1.62:1. When all are aggregated, the ratio is 1.89:1, suggesting that the characterization of voice may be more important in the construction of female characters. However, more interesting findings appear when we examine the contexts and distinguish specific voice qualities. The main categories that emerge are shown in Table 13.

	MBPC	FBPC
force / anger / control	17	6
sadness / weakness / fear	9	10
solemnity	3	0

Table 13. Main categories of voice qualities in the nineteenth century

For male characters, the most common quality mentioned, by a fair margin, involves force, anger, and/or exertion of control over others, as in examples 5 and 6.

- (5) *"Qui frappe" dit-il d'une voix furieuse.*
 "'Who's knocking?" he said in a furious voice.'
- (6) *"Entre!" cria-t-il d'une voix tonnante.*
 "'Come in!" he cried in a thundering voice.'

Among the female characters, on the other hand, the most commonly occurring qualities are in the area of showing sadness, weakness, or fear, or the speaker's emotional reaction to what they are experiencing, as in examples 7 and 8.

- (7) *"Sur quel vaisseau était monté ton père," demanda-t-elle d'une voix tremblante, ...*
 "'Which vessel did your father board," she asked in a trembling voice.'
- (8) *"Volés, volés" dit-elle d'une voix défaillante.*
 "'Stolen, stolen," she said in a failing voice.'

A quality that is gender-exclusive, though not large, is the category of solemnity, instantiated in three instances, all relating to male characters, with the adjectives *solennelle* 'solemn' in two cases and the adjective *sombre* 'somber' in a third.

For the twentieth century, comparing the data in Tables 11 and 12 suggests the same general tendency that has appeared elsewhere throughout our study: the nineteenth-century data show a rough parity between male and female characters, with a male:female ratio of 34:18, or 1.89:1, compared to the baseline ratio of 1.62:1, while the twentieth-century data show a divergent pattern, as the male:female ratio is 14:2, or 7:1, compared to the baseline ratio of 2:1. Further, two of the three five-grams are male-exclusive. Again, however, the small numbers call for skepticism; in particular, *s'écrit d'une voix* occurs in only one book, and its male-exclusive application may simply fall out from that fact. Further, a close examination of the qualities presents a tentatively different picture. Five of the sixteen usages can be confidently placed in

the force/anger/control category, and all of those are male body-part clusters. However, only two, or possibly three, are in the sadness/emotion category, and those are also male. The remaining five-grams evidence a range of qualities not falling neatly into categories established with the nineteenth-century data, such as mysteriousness, ambiguity, friendliness, coquettish or coaxing, etc. Both female body-part clusters are of this nature, as in one the quality is peacefulness, and in the other it is mysteriousness. If these findings could be taken as representative, a trend toward greater diversity of performance of femininity such as that noted for the English books might be supported, but the numbers are simply too small, and investigation with larger corpora is necessary.

CONCLUSIONS

Our findings highlight the need for more extensive French-language children's literature corpora to properly carry out this type of research. Nonetheless, we were able to draw some tentative conclusions from the available data. Our first research question was what methodological adaptations are necessary to study body part clusters in French. We found that Čermáková & Mahlberg's corpus-based five-gram methodology was easily transferrable with the necessary adjustments, in particular the change of focus toward context to account for the grammatical, rather than social, gender in French genitive pronouns. In fact, with a reasonable amount of time and effort, one can achieve a greater depth of detail with French data than by using the English-based methodology.

Our second research question had two parts, the first of which was whether body-part language in nineteenth- and twentieth-century French children's literature patterns like that in British literature. Our results parallel those of Čermáková & Mahlberg for the nineteenth century: differences between male and female characters are based on degree, not category, categorical usages tend to characterize males, and male-associated clusters lean more toward independently achieved actions, while female-associated clusters lean more toward interpersonal actions. In contrast to Čermáková & Mahlberg, however, we found a high degree of divergence between the nineteenth and twentieth centuries. A tentative exception appears with regard to quality of voice, which we discuss below, but the general pattern is that, to the extent that female-associated clusters occur, there is little variety. Even with the caveat that our data only covered the first half of the twentieth century and that our corpus was smaller, we find the degree of divergence surprising, given the well-known societal trend in Western Europe at the time that saw a greater visibility of women. Rather than hazarding an uninformed guess as to why such a difference was present, we will leave the matter for future research.

The second part of the question concerned what other tendencies the data might show. First, we found differences in the specific most-frequent five-grams. A prominent example was *de la tête aux pieds* 'from head to toe' / *des pieds à la tête* 'from toe to head,' which appeared 31 times total (equating to more occurrences than many of the clusters cited by Čermáková & Mahlberg, with a corpus 6.5 times larger than ours), and whose counterpart does not appear in the English-language analysis. The finer analysis of these expressions indicated that they were used for either psychological/emotional or physical character development. Each of these two categories further divided into two subcategories, which showed gender-specific tendencies. Notably, psychological/emotional conditions associated more with female characters, while physical conditions associated more with male characters, and the latter conditions were nonexistent in our twentieth-century data. This surprised us because we would have expected a

greater emphasis on the physical condition of female characters, due to societal tendencies to objectify women and girls.

Finally, our third question asked whether voice terminology also contributes to gendered construction of characters, as preliminary examination of the data suggested. Deeper analysis confirmed that the specific quality of voice does appear to be important, as a specification of a certain voice quality almost invariably accompanies mention of voice in frequently occurring five-grams. Further, the specific qualities referenced in the nineteenth century largely fall into defined categories, which assort by gender. Specifically, voice qualities associated with strength or force tend to describe male characters more frequently, while qualities associated with emotion, sadness, weakness, or fear tend to construct female gender. In the twentieth century, the data suggest that the distinction may lessen, in contrast to the general increase in divergence, but the paucity of data in that corpus discourages us from drawing any conclusions.

In closing, French children's literature clearly warrants more research, with a larger corpus. It seems safe to say that in the nineteenth century, French children's literature constructed a male-female binary akin to that observed by Čermáková & Mahlberg. On the other hand, the differences between their study and ours do seem to underscore the cultural dependence of clusters containing a body part and a gendered term. Investigation of the second half of the twentieth century, and indeed the twenty-first, will no doubt yield a more nuanced comparison.

Works Cited

- Adam, Helen, and Laurie J. Harper. "Gender Equity in Early Childhood Picture Books: A Cross-Cultural Study of Frequently Read Picture Books in Early Childhood Classrooms in Australia and the United States." *The Australian Educational Researcher*, vol. 50, no. 2, 2023, pp. 453-479.
- Butler, Judith. *Gender Trouble: Feminism and the Subversion of Gender Identity*. Routledge, 1990.
- Čermáková, Anna, and Michaela Mahlberg. "Gendered Body Language in Children's Literature Over Time." *Language and Literature*, vol. 31, no. 1, 2022, pp. 11-40.
- Chater, Nick, and Christopher D. Manning. "Probabilistic Models of Language Processing and Acquisition." *Trends in Cognitive Sciences*, vol. 10, no. 7, 2006, pp. 335-344.
- Filipović, Katarina. "Gender Representation in Children's Books: Case of an Early Childhood Setting." *Journal of Research in Childhood Education*, vol. 32, no. 3, 2018, pp. 310-325.
- Hamilton, Mykol C., et al. "Gender Stereotyping and Under-representation of Female Characters in 200 Popular Children's Picture Books: A Twenty-First Century Update." *Sex Roles*, vol. 55, 2006, pp. 757-765.
- Kok, Jodi LY, and Bruce Findlay. "An Exploration of Sex-Role Stereotyping in Australian Award-Winning Children's Picture Books." *The Australian Library Journal*, vol. 55, no. 3, 2006, pp. 248-261.
- Mahlberg, Michaela. *Corpus Stylistics and Dickens' Fiction*. Routledge, 2013.
- McCabe, Janice, et al. "Gender in Twentieth-Century Children's Books: Patterns of Disparity in Titles and Central Characters." *Gender & Society*, vol. 25, no. 2, 2011, pp. 197-226.
- Pownall, Madeleine, and Nathan Heflick. "Mr. Active and Little Miss Passive? The Transmission and Existence of Gender Stereotypes in Children's Books." *Sex Roles*, vol. 89, no. 11, 2023, pp. 758-773.