child and *children* in a corpus of American fiction:

Contrasting semantic preferences and their experiential motivations

John Newman

University of Alberta, Canada & Monash University, Australia

Abstract. This study investigates the nature of the semantic differences between collexemes of singular *child* vs. plural *children* in a corpus of American fiction. Collexemes in three specific syntactic slots are investigated: adjectives in the attributive position, nouns grammatically possessed by *child's* and *children's*, and present participial forms of verbs in the position immediately to the right of *child*/*children*. The approach relies on Distinctive Collexeme Analysis to establish statistically significant differences, revealing a surprising variety of semantic distinctions not usually acknowledged in discussions of the singularity vs. plurality of count nouns. Finally, an attempt is made to relate the key findings to broader cognitive and behavioural realities.

1. Introduction

In this study, I consider some lesser known properties associated with singularity and plurality in language, based on an extended discussion of the specific pair *child* and *children*. As part of the discussion, I suggest ways to relate the linguistic findings to broader behavioural and cognitive realities.

Attention will be focused on preferred patterns of co-occurrence of various lexical items, i.e., collocates, with singular *child* vs. plural *children* in some specific constructions. The term 'semantic preference' is an appropriate one to describe the kinds of co-occurrence patterns under investigation in the present study (Stubbs, 2001; Stubbs, 2016; Partington, 2004). An item is said to show a semantic preference when it co-occurs with "a class of words which share some semantic feature (such as words to do with 'medicine' or 'change')" (Stubbs, 2001: 88). In other words, the intention is not just to identify individual words that co-occur with *child*/*children*, but rather to identify classes of

co-occurring words and to characterize the semantic preferences that emerge in these classes.

The attention to semantic preferences is a natural consequence of a usage-based approach in which the context of use and frequency of use are key aspects of an analysis and can be traced back to Sinclair (1991), Stubbs (1995, 2001), Hoey (1991, 2008), Partington (1996), Biber, Conrad, and Reppen (1998: 21-54), Hunston (2002: 68-79), among others. It is now commonplace in corpus linguistics to investigate collocates, not just the collocates in some vaguely defined neighbourhood of a word, but also collocates occurring within a construction, in which case one speaks more specifically of *collexemes* occurring in one or more slots of a construction which can be subjected to a *collostructional analysis* (Stefanowitsch and Gries, 2003; Gries and Stefanowitsch, 2004). The present study is an example of such collostructional analysis centered on collexemes of *child* and *children* and the semantic preferences of these collexemes.

In non-usage-based approaches, the study of singular vs. plural nouns, or number in nouns more generally, has mainly focused on the nouns themselves without regard for collocational properties as may be revealed through corpus-based methods. In many studies the focus has been on the morphology associated with plural marking, whether the singular or plural form is marked by some additional morpheme added to the stem, irregularities in the formal expression of number in a noun as in double marking of the plural, etc. Such studies have helped to refine our understanding of the form-meaning relationship associated with grammatical number but shed no light on the context of use. Another line of research is typological, where the aim is to study the system of number distinctions (singular vs. dual vs. triadic etc.) and how these systems are realized in languages and how the systems vary across languages (cf. Corbett, 2000). For the most part, this research compares descriptions of number as found in published grammars and as such has little to say about contextual preferences in how particular examples of singular and plural nouns are used in discourse. In the more formal semantic tradition, there has been a focus on how the notion of plurality is to be precisely interpreted, typically in set-theoretical and symbolic-logical terms. Nouwen's (2016) chapter on plurality in *The Cambridge Handbook of Formal Semantics* offers a contemporary overview of the main themes in this tradition, such as: whether the interpretation of

plurality is distributive as in *X and Y were wounded* or non-distributive as in *X and Y carried the piano up the stairs*; possible ways of understanding examples such as *six frogs swallowed twelve ladybirds* and *three boys wrote an essay* and the set-theoretical representations of how the agents and actions are connected; a special interest in the quantifiers *each* and *all* and the interpretative issues which arise in sentences like *all bankers wear suits*. These kinds of (fictional) sentences are intriguing in their own ways but they constitute an extraordinarily narrow slice of the full range of use of plural nouns and have nothing to say about the semantic preferences of a plural form in contradistinction to a singular form. Part of the motivation for the present study is to counter this relative neglect of collocational properties of singularity vs. plurality in the traditional linguistic linguistic/philosophical approaches.

The present study also builds on research into collocational properties of the inflected forms of a lemma, as opposed to analyzing collocational properties at the lemma level only. Stubbs (2016:112) cites early examples of an interest in such differences in Sinclair (1991: 53–64, 154) in a discussion of the lemma YIELD and Stubbs' (2001: 28) own investigation of the lemma SEEK. Sinclair had found, for example, that the 'give forth or supply' sense of YIELD is most strongly associated with the inflected form *yields* as opposed to the other inflected forms. Other studies of inflectional differences include the discussion in Newman (2010: 87-89) about differences in collocates of the positive and superlative adjectival forms *slight* and *slightest*, where the former occurs preferentially with nouns with the meaning of 'change, variation' while the latter occurs with nouns with meanings related to 'cognition, perception, intention'. Newman and Rice (2004) report on how the inflectional differences between the *–ing* and past tense forms in the pairs *sitting and.../sat and..., standing and.../stood and, lying and.../lay and...* profoundly influence the range of following verbal collocates. Deignan (2006) discusses the role of inflectional differences in studies of metaphor in general, as well as reporting on specific differences in metaphorical use of singular vs. plural forms such as *flame* vs. *flames* and *rock* vs. *rocks*. Studies such as these offer glimpses into the semantics of inflectional differences in English and taken together make a good case for the value of studies into the semantics of inflectional differences.

Apart from studies involving very specific lexical items as the focus of interest, two studies explore collocational/collostructional behaviour of inflected word forms in more computational ways, reporting results across large numbers of lexical items. The first of these, Gries (2011), investigates consequences of carrying out collostructional analysis based on different levels of granularity, inflected forms vs. lemmas, as well as single genre vs. multiple genres and argues that the more comprehensive approach may often be more revealing. In one of the case studies in Gries (2011), the author considers similarities and differences in collostructional behaviour of inflected forms of verbs (infinitives, ing-participle, past participle, past tense, present tense) in the English ditransitive construction compared with the corresponding lemmas. He finds, as one result among many, that the two verbs most strongly attracted to the ditransitive construction in the lemma analysis, *give* and *tell*, are also the two verbs most strongly attracted to the construction at the inflected word form level. His final conclusion is that the overall preference for particular semantic classes of verbs in the construction, as determined by collostructional analyses, is the same in both analyses. The second of these studies is Katz and Zamparelli (2012) which reports on a vector space analysis of the singular vs. plural distinction in 2,114 nouns in a POS-tagged corpus that occur with both singular and plural POS-tags forms. The window of the context for collecting the collocates of each noun was the four previous and four subsequent content words. The authors find that their methods "clearly show that there is a large proportion of nouns whose distribution in the singular and the plural differs, in terms of neighbouring content words." (Katz and Zamparelli ms: 43). The authors' particular focus is on mass nouns and the 'meaning-shifting' that arises when mass nouns appear as plural nouns (*wine* vs. *wines*) where one might in fact expect singular vs. plural distinctions among collocates. Apart from the meaning-shifting involving mass nouns, Katz and Zamparelli (2012: 378) observe: "We have also uncovered a wide range of other cases in which the singular and plural forms of a noun have different distributions, itself an interesting result (and one frequently ignored by vector space models)." The two sets of findings from Gries (2011) and Katz and Zamparelli (2012) are not that easy to compare, since the methods and focus of research are quite different, even if both are concerned with differences between inflected word forms. To the extent one can compare them, though, the findings point in different

directions, with the former arguing against inflected word forms as a locus of special interest and the latter suggesting the opposite.

The choice of *child/children* as the focus of interest has, it should be said, no compelling linguistic or cognitive rationale. Kinship terms in languages do display a range of properties of special interest to linguists and anthropologists, e.g., they may show special morphosyntactic properties, and they often participate in various semantic shifts resulting in a relatively high degree of polysemy. The latter inclines one to expect that the collocational behaviour of *child/children* may indeed lead to richer and more interesting results than, say, a consideration of *giraffe*, *photocopier*, or *hypotenuse* would, and these considerations were the main motivation for the choice of *child/children* as the focus. The plural form *children* is noteworthy and unique in English for its particular kind of plural marking which, historically, amounts to 'double plural marking', the *-r-* and the *–en* of *-ren* being historically plural markers of nouns in their own right (see Tiersma 1982 for further discussion of this phenomenon). But there is no particular reason to expect the collocational behaviour or, more generally the semantics, of *children* to be of special interest just because the morphology of the form is unique.

The present study aims to discover co-occurrence patterns that would not be very obvious, or obvious at all, simply from arm-chair reflection on the words *child* and *children*. Some co-occurrence patterns are quite obviously determined by the singular vs. plural distinction in the nouns, such as the occurrence of *a/one* with singular nouns only, *many/few/a few* or the cardinal numbers *two, three, four* etc. with plural nouns only, singular nouns as heads of subject NPs occur with verbs in the singular forms, etc.[1] These co-occurrence patterns are a systematic part of English grammar and are familiar to linguists and language users alike, without the need for discovery or substantiation by corpus linguists.[2] The focus in this study will be, rather, on three construction types where there is no obvious *a priori* reason to expect very different co-occurrence patterns: (a) adjective + *child/children*, (b) *child's/children's* + noun, and (c) *child/children* + present

---

[1] One should note that *two* can occur with *child* as a hyphenated form *two-child* in compound nouns such as *two-child family*.

[2] Where there is a stylistic or dialectic variation in verb agreement, then of course a corpus-based study of such variation is called for, e.g., the absence of 3rd singular *–s* marking on verbs (*she walk, she have money*), or the use of one simple past tense form of the verb *be*, either *was* or *were*, used with either singular or plural subjects.

participle of a verb. The constructions thus lead to collexemes belonging to three major parts of speech in English: adjectives, nouns, verbs. The intention is not to arrive at conclusions about collostructional behaviour with all nouns in English which would clearly be unwarranted. Rather, the intention, as in the other studies of collocational behaviour referred to above, is to take one example of the noun class and explore the nature and extent of the collocational differences. To the extent that the contrasting semantic preferences found with *child/children* are novel and can be motivated in principled ways, then this study may be taken as an incentive for further exploration of other members of the noun class in English.

Section 2 introduces the corpus and the method of collostructional analysis. Section 3 presents the main findings concerning the collexemes as revealed by the collostructional analysis, identifying the semantic preferences in whole classes of collexemes. In Section 4 I consider, more controversially, how the more important semantic preferences identified in the collostructional analysis reflect various behavioural and cognitive realities. In seeking out what one could call broadly 'experiential' motivation for the significant collostructional patterning, I align myself with the many researchers in the cognitive linguistics field whose preference is to explore the motivations, as much as possible, for the way we use words, leading to a more reduced role for arbitrariness in linguistic explanation (cf. Radden and Panther, 2004; Panther and Radden, 2011). Section 5 is the Conclusion.

2. Data and method

The data will be taken from COCA, the Corpus of Contemporary American English (Davies, 2008-), relying on the online search tool in COCA to arrive at the frequencies of the forms. COCA includes five genres – spoken, fiction, magazine, newspaper, academic – and a preliminary exploration of the most frequent collexemes of *child/children* revealed considerable variation between genres. As interesting as a cross-genre study of the phenomena may be, the decision was made to base the present study on a specific genre, namely the General (Books) subcategory of the fiction genre, consisting of novels published in the period 1990-2017, avoiding issues relating to variations between genres.

At the time this research was carried out (August – December 2019), this particular sub-corpus consisted of 24.5 million words, according to the online information at the COCA site. This sub-corpus is sufficiently large to yield substantive results for all three construction types and yet not so large as to make it impractical to double-check all concordance lines for accuracy and correct interpretation of forms in context, as was done routinely for the corpus results reported on in Section 3. Results from the present study cannot be taken to be representative of all genres, though a larger study comparing genres would of course be valuable. Davies and Gardner (2010), in fact, summarize key collocational properties of words (at the lemma level) across all five main genre types of COCA. I return to a brief comparison of their results for the lemma *child* with results from the present study in Section 3.4. Henceforth, I will refer to the General (Books) sub-corpus of COCA simply as 'the corpus'.

Even a cursory comparison of results from searching the corpus for the prenominal adjective collexemes of *child/children* showed that there are many shared collocates: 347 *only child* vs. 33 *only children,* 124 (*an*)*other child* vs. 227 *other children*, 38 *young child* vs. 68 *young children,* 37 *poor child* vs. 12 *poor children* etc. With such data, it is by no means obvious which adjectives are attracted more to *child* vs. *children* when overall frequencies are taken into account. The type of collostructional analysis that targets this kind of data, where the focus is on the competition for collexemes between two similar constructions (here, a *child* construction and a *children* construction) is Distinctive Collexeme Analysis (DCA, cf. Gries and Stefanowitsch, 2004; Gries, 2012).[3] Consequently, DCA was carried out on the three constructions of interest using the Collostructions package in R (R Core Team, 2019; Flach, 2017). The contingency table underlying the calculation of the collostructional strength of *poor + child/children* within the corpus is shown in detail in Table 1 in order to illustrate the DCA method. 'Other adjectives' in Table 1 refers to the sum of the frequencies of other adjectives occurring with either *child* (in column 2) or *children* (in column 3), and 'All

---

[3] Gries (2019), an important and far-reaching critical review of collostructional methods, argues for a greater appreciation for, and more reporting of, a variety of measures in corpus linguistics. This includes reporting multiple measures relating to word associations, rather than reporting associations as single scores, as is the case with DCA using Log Likelihood scores, as in this study. The need to consider a range of corpus-linguistics measures would appear to be most urgent when the goal is to compare the corpus-linguistics results with results from other methods, e.g., experimental methods. This is not the goal of the present study and reporting the collostructional strength scores in the conventional way seems to be still the most convenient way to proceed.

adjectives' refers to the sum of the frequencies of all adjectives occurring with either *child* or *children*. Determining expected vs. observed frequencies of collocates is integral to DCA and the calculation of the expected frequency of *poor* + *child* in the contingency table of Table 1 is shown in (1a-b). The expected frequency of *poor* is 28.5 with *child*. So, the observed frequency of *poor* + *child* (37) is more than what one would expect and one can say that *poor* is preferentially attracted to the adjective + *child* construction. The collostructional strength is determined by the Log Likelihood which in this case is 6.5891, df = 1, $p = 0.01026$, i.e. < 0.05 (*).[4]

Table 1. Contingency table showing observed values used for calculation of collostructional strength of *poor* + *child* vs. *poor* + *children* in the corpus.

	Observed Frequency with *child*	Observed Frequency with *children*	Observed Frequencies of adjectives in both constructions
poor	37	12	49
Other adjectives	2,046	1,490	3,536
All adjectives	2,083	1,502	3,585

(1a) Overall proportion of adjectives with *child*

= Frequency of adjectives with *child* / Frequency of all adjectives with *child* and *children*

= 2083/3585

= 0.581

(1b) Expected frequency of the adjective *poor* with *child*

= Overall proportion of adjectives with *child* x frequency of *poor* in both constructions

= 0.581 x 49

= 28.5

---

[4] In all tables reporting significance levels, the significance level is indicated by asterisks thus: ***** = significant at $p < .00001$, **** = significant at $p < .0001$, *** at $p < .001$, ** at $p < .01$, * at $p < .05$.

3 Analysis

3.1 *Adjective* + child/children

A search for adjective + *child/children* sequences in the corpus returns 1,986 hits for adjective + *child* and 1,524 hits for adjective + *children*. Some editing of these results (and other results in the following sections) is necessary to improve precision and recall. In terms of precision, there are indeed various categories of false positives. All instances of *still* + *children* result from a part-of-speech tagging error whereby the adverb *still* has been assigned an adjective tag when it actually functions as a temporal adverb, as in (2a). Instances of *like* as in (2b) were also removed from the results since *like* is functioning as a comparative relational adverbial rather than as an adjective qualifying *child*. In some cases, the adjective is modifying a possessed noun, as in (2c-d) or a larger compound noun, as in (2e), rather than *child/children* and these instances were removed from the final results.

(2) a. From a certain point of view the sisters looked like women, tall and beautiful and poised, but they were <u>still children </u>in many ways, the younger girls especially. (Alice Hoffman, *The Story Sisters,* 2009)

   b. I clutched my daughter's fingers tighter, feeling more <u>like child</u> than mother. (Gina Holmes, *Crossing Oceans*, 2010)

   c. She shot a glance across the room at the two men, slumped on the <u>tiny children's desks</u>. (Ira Sher, *Gentlemen of Space*, 2003)

   d. …it would have made that whole dread business <u>mere child's play</u>. (Leanna Renee Hieber, *Eterna and Omega*, 2016)

   e. "I'm Barb Hunter, the <u>new Child Protection Specialist</u> for the nursery unit." (Linda Raymond, *Rocking the Babies*, 1994)

As far as recall is concerned, one should note that a search on the adjective category before *child*/*children* does not return any instance of *another*. In the CLAWS tagging system relied on by COCA, *another* is tagged as a determiner (DD1) while *other* is tagged

as an adjective (JJ). To the extent that *another* is considered an orthographic peculiarity of the sequence *an other*, then the frequencies of *other* and *another* should be combined into one total for *(an)other* before *child*, which gives a frequency 124 for the combined category instead of 22 occurrences of just *other*. Despite some distributional idiosyncrasies of *another* vs. *other*, for the present purposes it seems more appropriate to proceed in this way and all instances of *another child* will be regarded as 'true negatives' and added to the results from the online search in COCA. There is no instance of the sequence *another children* in the corpus.

The revised frequencies after these corrections are 2,083 (adjective + *child*) and 1,502 (adjective + *children*). Clearly, *child* attracts more adjectives than one would expect, given that *children* is more frequent than *child* in the corpus. It points to a greater descriptive elaboration in the discourse about a child compared with children.

I will present the analysis of the singular construction followed by the plural construction (see tables in Appendices 1 and 2).[5] In all the DCA tables, COLLEX refers to the collexeme, O and E refer to the observed and expected frequencies of the collexeme, ASSOC refers to the construction that the collexeme is more attracted to, COLL.STR refers to the collostructional strength as measured by Log Likelihood, SIGNIF refers to significance level as explained in Footnote 1, and Shared refers to whether or not the collexeme is shared between the two constructions. Only collexemes occurring with frequency >5 for the combined *child* and *children* frequencies are displayed in the case of larger numbers of results (Appendices 1-4).

(3) summarize the main results from the DCA in Appendices 1 and 2, arranged into smaller groups for convenience of exposition. The summaries in (3a) and (3b) have been constructed to illustrate larger trends in the results and exclude a few miscellaneous adjectives.

(3) a. Adjectives more attracted to *child*:
        only 'not having siblings', *single, Holy*
        middle, oldest, eldest, youngest
        dear, precious, poor

---

[5] I list only the results that are statistically significant in Appendices 1-4, but include all results for the smaller amount of data in Appendix 5 (the present participial collexemes).

petulant, sensitive, frightened, spoiled, wilful

b. Adjectives more attracted to *children*:
other
older, young, younger
deaf, Jewish, white, school-age, local
little, small

The adjectives in the group {*only, single, Holy*} all have a semantic connection to some 'unique' condition. *Holy Child* in its most literal sense refers to the unique infant Jesus of Christianity, though the combination is used in the corpus to refer to some other individual in various extended senses of (uncapitalized) *holy*. The *only child/children* combinations can be analyzed as compound nouns in some cases, based on the unpredictable semantics of the combination meaning 'single (i.e., unique within a family unit) offspring' and the location of stress on the adjective rather than the noun in these forms.[6] Note that *child* and *children* both occur with *only* in this sense, even within the same excerpt as in example (4a). Example (4b) illustrates the distinct 'restrictive' use of *only*, in this example restricting the referents of *children* to those who are Caroline's age. As it happens, it is only the 'single offspring' sense that is found with *only child* in the corpus, whereas *only children* occurs with both senses as illustrated in (4a) and (4b). *Other* behaves in the opposite way to *only*, being attracted more to the plural *children*.

(4) a. "But you were an <u>only child</u>. <u>Only children</u> are usually very close to their parents." (Barbara Taylor Bradford, *Her Own Rules*, 1996)
b. These are only boys, <u>only children</u> Caroline's age, and there is no reason, just because they are black, that I should be afraid of them. (Susan Rebecca White, *Bound South,* 2009)

Somewhat related semantically to the above is the group {*middle, oldest, eldest, youngest*}, consisting mostly of age-related superlative adjectival forms. These adjectives, too, prototypically identify a single entity among many, in this case by virtue of the

---

[6] I am grateful to one of the reviewers of this article for pointing this out to me.

position within a scale or rank. The example in (5a) exemplifies this property in that the eldest child is being directly contrasted with all the children younger than her. In (5b), *eldest children* refers to the set of children who are each the single eldest child within a family, rather than referring to multiple children within a family. This group is in direct contrast to the group {*older, young, younger*}, consisting of age-related plain ('positive') and comparative adjectival forms which are more strongly attracted to the plural.

(5)   a.   Mary, the <u>eldest child</u> at twelve, kept the younger ones occupied on a quilt set up in the corner. (Pam Hillman, *Stealing Jake,* 2011)

     b.   So new to widowhood, she had already discovered its power - unless she was just doing what <u>eldest children</u> always did, bulldozing over her sibling's needs. (Katherine Shonk, *Happy Now?,* 2010)

The remaining adjectives attracted to *child* fall into a few semantic groups. The group {*dear, precious, poor*} includes those adjectives that mainly serve to express a strong degree of empathy, evoking love, endearment, fondness, sympathy etc. All three adjectives are polysemous. *Dear* and *precious* can mean 'expensive' but the empathetic sense dominates in the data and is the only sense that appears with *child* or *children. Poor* can have the 'financially impoverished' sense with *children* in one or two cases of the twelve instances of *poor children*. The examples of *precious* and *poor* in (6a) and (6b) illustrate these empathetic uses. Yet another group of adjectives attracted to *child* is the group {*petulant, sensitive, frightened, spoiled, wilful*} describing human behaviours relating mainly to the personality or state of mind of a child.

(6)   a.   DNA samples confirmed that the <u>precious child </u>I'd buried two years ago wasn't mine, and that Andrea Hayley Lockhart was actually my biological child. (Debbie Fuller Thomas, *Tuesday Night at the Blue Moon*, 2008)

     b.   "My <u>poor child</u>," said Sally, "if I had known, I would have been a lot easier on you… (Alexandra Ripley, *Scarlett*, 1991)

The remaining adjectives attracted to *children* fall into a couple of groups: {*deaf, Jewish, white, school-age, local*} and {*little, small*}. The former group consists of adjectives that describe miscellaneous kinds of classifications other than psychological. The latter group consists of size-related adjectives.

3.2. child's/children's + *N*

In total, there are 588 instances of *child's* + N and 504 instances of *children's* + N, so again more instances of the *child's* + N construction than one would expect given that there are more instances of *children* than *child* in the corpus. Some of the nouns (ten in all) occurring immediately after *child's* and *children's* were the first element of a noun compound necessitating some manual correction to the set of possessed nouns in the results, e.g., *child's play scene*, *children's book illustrator*, *children's book catalogs*, *children's book character*.

(7a) and (7b) summarize the main results from Appendices 3 and 4 regarding *child's/children's* + N. *Child's play* is a fixed expression with its own meaning of 'human behaviour demonstrating skill in responding to a supposedly challenging task' only ever occurring in this form and is excluded from the more detailed discussion below.

(7) a. Nouns more attracted to *child's* + N:
    face, hand, voice, body, forehead, hair, cry, eyes
    game, doll, toy
    life, death, name, mother, father

   b. Nouns more attracted to *children's* + N:
    book, books, toys, clothes
    hospital, clinic, section, ward
    voices, faces,
    lives, children

The group {*face, hand, voice, body, forehead, hair, cry, eyes*} consists of body-related terms, here including references to sounds emitted by humans, *voice* and *cry*. With some of these terms, one would hardly expect them, in light of human anatomy, to be the possessed noun occurring with a plural *children's* possessor. So, for example, there are no instances of *children's face/hand/body/forehead*, just as there are no instances of *child's faces/bodies/foreheads* in the corpus. On the other hand, one does find *children's hands/voices/faces/cries* in the corpus (though not the biologically possible *children's bodies/foreheads*). Miscellaneous inalienably possessed items other than body-parts occur with both *child's* and *children's* in Appendices 3 and 4, but this category is more common and more significant with *child's* than with *children's*. Compare, for example, the corpus frequencies in pairs such as: *child's life* (22) vs. *children's lives* (6), *child's death* (10) vs. *children's deaths* (1).

The category {*hospital, clinic, section, ward*}, referring to buildings or parts of buildings, is attracted to *children's* in strong contrast to the body-parts and other inalienable entities.

Human artefacts intended for the use of a child or children appear in both (7a) and (7b) above. {*game, doll, toy*} are more attracted to *child's* + N while {*book, books, toys, clothes*} are more attracted to *children's* + N. Note that *child's game* can be used like *child's play* referring to adult behaviour. Cf. (8a) where the reference is to an actual game played by children and (b) the 'child's play' sense.

(8) a. Then it, could disappear, as our fingers crossed, interlocking as in
the <u>child's game</u> of church and steeple,… (Charlie Smith, *The Lives of the Dead*, 1990)

 b. He'd turned the court proceeding into a <u>child's game</u>. (Jay Brandon, *Angel of Death*, 1999)

3.3 child/children + *Ving*

Some editing of the returns from COCA was necessary, as for the other constructions. *Rearing* is tagged as an *-ing* verb in the corpus compound, but the combination *child rearing* is an instance of a compound noun rather than a noun qualified by Ving. Consequently all instances of *child rearing* were removed. Also, in the case of *chasing*, two instances were found to be counted twice by COCA and were corrected. After editing of the returns, there were 168 instances of *child* + Ving and 271 instances of *children* +Ving. The verb collexemes are very sparsely distributed in each of these constructions: 83% (140/168) of the verb collexeme tokens with *child*, and 86% (233/271) of the verb collexeme tokens with *children*, occur just once in the construction. The collexemes that do occur multiple times in a construction occur with low frequency, as can be seen in Appendix 5 and only a handful of verbs manage to reach statistical significance in the data. Nevertheless, at least one larger pattern can be seen in the significant results.

Of the three verbs that have a significant association with *children*, two, interestingly, refer to motion: *running* and *coming*. *Running* is particularly noteworthy as it has the highest collostructional strength and is the only collocate which has a significant association at $p < .01$. The combination *children running* is typically used in the corpus to refer to relatively uncontrolled, unsupervised running, as in *children running free*, *children running about, children running about in packs, a million children running all over the place like wild Indians*, etc. The combination *child playing*, on the other hand, may refer to either translatory motion (*a child playing tag*) or non-translatory motion of the body (*a child playing the violin*) in the corpus. Although just these two motion verbs show up as associated significantly with *children* in the DCA, it was decided to probe a little further to explore whether motion shows a preferential association with *children* in some other contexts. To this end, a further search was conducted in the corpus targeting all lexical verb (as opposed to auxiliary verb) collocates within a span of two words to the right of *child* and *children*. This span does not correspond to any well-defined construction; rather, it includes miscellaneous construction types such as subject-verb constructions (*if the child died*), past participial phrases (*a child born from evil*), relative clause constructions (*a child who needed her help, each child I felt for a pulse*), conjoined structures (*take my child and give her to her autocratic and loveless father*), etc. Still, exploring the most frequent verbal collocates of *child* and *children* not restricted to a

particular construction type may provide some further insight, and brings with it the benefit of working higher frequencies.

The top results of the search for these verbal collocates of *child* and *children* are shown in Table 2.[7] I have chosen to simply list the ten most frequently occurring collocates for *child* and *children*. The top ten collocates with *child* are in fact all the collocates with frequency >11 and the top ten collocates with *children* are all the collocates with frequency >21, providing a more robust set of results than relying only on the Ving form. Collocates which involve (translatory) motion have been highlighted in the table. It is striking that 8/10 of the top collocates with *children* involve motion, whereas only 2/10 of the top collocates of *child* do.[8] I have chosen to include *playing* and *played* as motion verbs even if it is sometimes translatory and sometimes not (the more contemporary idea of playing a computer game in a seated position was not evident in the data). Even without investigating in detail each of the construction types underlying the collocates in Table 2, the albeit rather crude results from working with such a broad sweep of constructional types lend further weight to the finding from the DCA that motion verbs, especially translatory motion verbs, are more strongly attracted to *children* compared with *child*.

Table 2. Most frequent lexical verb collocates (0-R2) of *child* and *children*. Motion verbs are highlighted.

Rank	*child* collocates	Frequency	*children* collocates	Frequency
1	*born*	41	*playing*	47
2	*come*	19	*came*	39
3	*playing*	18	*come*	36
4	*died*	17	*played*	31
5	*knew*	16	*went*	29

---

[7] Comparable results are found with the pair *kid* and *kids* in the corpus.
[8] *Get* can involve some physical motion but the instances are too few to justify marking *get* as a motion verb in Table 2. In the corpus results, *get* is used to refer to physical motion with *child* in just 3/15 cases where it appears in the 0-R2 range of *child*: *the child would get in my closet*, *he saw a woman and a child get out of the car*, and *it's like watching your child get on the school bus*. *Come* and *go* and their various inflected forms can have non-motion uses (*she went red in the face*, *he went crazy* etc.) but these are not frequent in the search results.

6	*said*	16	*running*	24
7	*get*	15	*born*	23
8	*lost*	14	*gone*	23
9	*needed*	12	*grown*	23
10	*seemed*	12	*ran*	22

3.4. *Methodological note: maximally constrained vs. 'broad sweep' methods*

Throughout this study I have followed what may be called a maximally constrained approach: studying collexemes at the level of *inflected word forms* rather than lemmas, finding patterns within *specific construction types* rather than across construction types, and choosing to explore *one specific genre* of fiction rather than a mix of genres.

Making different choices for all these parameters may lead to different results from those reported on here, and a comparison with Davies and Gardner's (2010) method and results brings home how different the findings can be when different procedural choices are made. Davies and Gardner (2010) is a unique corpus-based dictionary of collocates of English, listing the most significant associations of collocates for each head word of the dictionary. Their approach is the exact opposite of the approach adopted here and can be described as 'broad sweep', working with lemmas not inflected word forms, identifying significant collocates in the window L4-R4 rather than construction-specific collexemes, and using the whole of COCA for data rather than a specific genre. The adjectival collocates of CHILD as listed in their dictionary (which is to say in order of decreasing significance) are shown in (9).

(9)   *young, gifted, poor, foster, healthy, educational, elementary, pregnant, emotional, unborn*

Most of these adjectives play no role in the results from the DCA in the present study. Partly, this is a consequence of Davies and Gardner's decision to rely on the whole COCA with many of these collocates appearing so dominant because of high frequencies in

academic and newspaper genres. The reliance by Davies and Gardner on the whole COCA corpus prevents miscellaneous and less frequent patterns of co-occurence in the sub-corpus of General (Books) fiction from making an appearance in the dictionary, such as the pattern of empathetic adjectives like *dear* associating with *child*. Of the adjectives in (9), only *young* and *poor* appear in the DCA results in the Appendices. And for these adjectives, the lemma-based approach of Davies and Gardner cannot possibly identify the contrasting associations for the inflected word forms identified by the DCA: *young* has a significant association with *children* and not *child*, whereas *poor* has a significant association with *child* and not *children* (see Appendices 1 and 2). There is no question that Davies and Gardner (2010) is a unique and valuable contribution to the study of English language usage, but its value lies in its breadth and overview and does not contribute to our knowledge about construction-specific, inflection-specific patterns, as exemplified by the present study.

4. Experiential motivations

In this section I select some of the key semantic preferences reported on in Section 3 for further discussion. My interest lies mainly in seeking experiential motivations for the contrasting semantic preferences of *child* and *children*. Broadly speaking, this means relating the linguistic findings to aspects of human behaviour and cognition apart from language.

4.1 *Adjective* + child/children

The first point of interest concerns the contrast between the adjective collexemes of *child* referring to psychological categorization (*petulant, sensitive, frightened, spoiled, wilful*) and the adjective collocates of *children* referring to non-psychological, non-behavioural categories (*deaf, Jewish, white, school-age, local, little, small*). The *child* collocates reflect judgements that are based on how a child behaves in certain situations, what a child might be feeling, the child's inner life etc., which is absolutely not the case with the *children* collocates. I have no convincing explanation for this result but it is intriguing. The

alignment of psychological vs. non-psychological adjectives with singular vs. plural marking in the accompanying nouns is not something that has been remarked upon or studied to the best of my knowledge. The alignment seems to suggest that psychological attributes are primarily conceived of as individual attributes, whereas the non-psychological adjectives are conceived of more naturally as group attributes, or at least not strongly individualistically. Obviously, one can speak of *petulant children*, just as one can refer to a *local child*, but the DCA shows these combinations are not the preferred choices, probabilistically.

A second point concerns endearment terms that are found preferentially with *child* (*dear, precious, poor*), where the speaker (narrator or character in a novel) is expressing a high degree of empathy for a child. This result is arguably a linguistic instantiation of a larger human behavioural phenomenon concerning the different ways in which we behave towards or reflect upon *individual* humans as opposed to a *collective* of humans. The psychological literature relating to empathy, which is substantial, is relevant in this regard, in so far as this literature understands empathy almost always as a state or feeling directed primarily towards others, specifically other individuals, rather than a collective. Cuff, Brown, Taylor and Howat (2016) conveniently provide an overview of definitions of empathy in the psychological literature and it is clear from these definitions that it is a single other individual person who is typically understood as the target of empathy. Hence, the references to 'the other person involved', 'the other person's mental state', 'an other oriented emotional response', 'the client', 'another's feelings', 'another person's condition', etc. in the definitions summarized in Cuff et al. (2016: 146-147). Main, Walle, Kho and Halpern (2017: 360) argue specifically for the strongly ('inherently') interpersonal nature of empathy, typically two individuals: "[…] in real life empathy is an interactive social process dependent upon *both individuals* for adaptive functioning" (my italics). The collexemes expressing empathy with *child* may in part be relatable to this larger phenomenon of empathy being most typically understood as a person-to-person form of behavior. One can, of course, feel empathy towards a group of individuals or a collective and *poor children* is indeed used in the empathetic sense in the corpus, but it is the singular use *poor child* that is significantly overused.

A third point of interest concerns the use of certain adjectives that arguably involve a unique entity as a prototype. *Only* and *single* are the most obvious examples as well as *holy* in *Holy Child*. One could argue, too, that the superlative form of an adjective, often characterized as referring to the highest degree of a quality in contrast to all other members of the category, is prototypically the single element at the high end of a scale, as seen in *oldest, eldest,* and *youngest. Middle* would seem to be allow more vagueness in its exact reference, e.g., *the middle of a circle* suggests a vague area within a circle rather than a unique point, but in the context of a family, the term does suggest the one child situated between older and younger siblings.

4.2 child's/children's + *N*

With *child's*, one finds preferences for collexemes that fall into the category of inalienable possession: terms relating to body-parts (*face, hand, voice, body, forehead, hair, eyes*) or bodily characteristics (*life, death, voices, cry*), kinship terms (*mother*, *father*), and personal names (*name*). These stand in stark contrast to the terms for buildings and parts of buildings with *children's*: *hospital, clinic, section, ward* which are preferred collexemes for *children's*. With *child's*, it is the human body and the human family, as traditionally understood, that underlie most of these preferences. With *children's*, it is buildings and parts of buildings that are preferred, though clearly still related to human goals and needs, specifically bodily needs (*hospital*, *clinic*, *ward*). This contrast between inalienable possession (with *child's*) and a specific type of alienable possession (with *children's*) is not something that is usually acknowledged as being associated with the singular vs. plural noun distinction in English. One can indeed refer to both *child's face* and *children's faces*, *child's hand* and *children's hands, child's mother* and *children's mothers, child's name* and *children's names,* etc. But the DCA has shown the greater likelihood of these expressions in the singular rather than the plural.

Underlying these preferences are two main cognitive models which I believe throw light on these findings (cf. Lakoff, 1987: 12-13, 68-90). The first relates to the primary

cognitive model relevant to understanding body-part terms, which is THE HUMAN BODY.[9] Of course, the human body (understood as a single entity) and human bodies (understood as multiple entities) are key concepts in cognitive linguistics and cognition more generally. These concepts are the foundation of the notion of embodiment in human cognition. Both the single entity and multiple entity concepts – uniplex and multiplex in Talmy's (2003: 55-61) approach – are relevant. The single entity concept of the human body is most relevant to understanding experiential realities involving touch, balance, front vs. back, centre vs. periphery etc. On the other hand, other experiences are grounded in personal interaction between human bodies, such as speaking, loving, giving, kissing, hugging, etc. Body-parts are most naturally understood as parts of a *single* human body. A hand, for example, is a part of a human body, contained entirely within one human body. A hand is not something that belongs simultaneously to multiple human bodies. This is a trivial observation in a way, but worth stating explicitly given the casualness about how both *the body* and *bodies* feature in the embodiment literature. A second cognitive model which is relevant here is the concept of FAMILY, as traditionally understood, consisting in its core of a mother and father and one or more children. Here, too, it is a single entity 'mother' and a single entity 'father' that are part of this (traditional) model. There are, of course, many cultural variations on the boundaries of what constitutes the FAMILY cognitive model, e.g., single-parent family, extended family, blended family, gay-parents family, etc. However, a mother and a father are the prerequisite for a family in the traditional model which is the one most relevant to understanding past language usage. The singularity of a father and a mother in the family cognitive model and the singularity of the human body in the human body cognitive model motivate, I argue, the attraction of the singular *child's* to the body-part and kinship terms. *Child* provides a more immediate and more natural 'point of access' to concepts relating to body-parts and kinship than does *children* (cf. Langacker, 1987: 163-164).

4.3 child/children + *Ving*

---

[9] For more detailed discussion of the literature relating to THE BODY cognitive model, see Gallager (2005), Gallager (2005: 24) distinguishes between *body image* as "a system of perceptions, attitudes, and beliefs pertaining to one's own body" and *body schema* as "a system of sensory-motor capacities that function without awareness or the necessity of perceptual monitoring". The body cognitive model relevant to the present discussion corresponds more to Gallager's concept of body image.

The data relating to lexical verbs is relatively sparse and the findings from this part of the study are suggestive rather than conclusive. The main finding to report is that *children* attracts motion verbs, especially translatory motion verbs, more than *child* does. A connection between *children* and motion verbs more generally is also suggested by the additional frequency data relating to all lexical verbs in the 0-R2 range. Furthermore, there is a tendency for the motion to be relatively unrestrained, disorderly, even chaotic, typified by expressions involving *running* such as *children running free, children running about*, *children running all over the place* etc. Obviously, children can congregate without engaging in this kind of behaviour. Children can sit as a group and remain relatively stationary – the image of children sitting at desks in a classroom springs to mind. But the DCA points to a preference for the plural *children* to be associated with running around, not sitting and remaining in one place. I am reluctant to read too much into this finding, given the relatively low frequencies of the collocates, but the finding, about motion verbs generally and *running* in particular, invites further research on this semantic connection.

I am only able to offer rather simplistic and largely unscientific comments relating to any specific behavioural association connecting children (as a collective, as opposed to an individual child) and the act of running around. Groups of children can often be seen enjoying themselves running around together in a playground, for example, but I can't confidently say this is a more common reality than just seeing an individual child playing in a playground. For a more informed comment, one could cite here a remark by the sociologist William Corsaro reporting on a study involving multiple children in an open space in a school setting, who makes the general point about children's running behaviour: "Children like to run, they like to move around. For young children running, jumping and laughing are in many ways equivalent to talk (or conversation) among older children and adults." Corsaro (1986: 233). It seems intuitively obvious that the more children congregate together, the more opportunity there is for them to engage in the kinds of activities described by Corsaro, which includes running. The more children who are present, the more scope there is for interactive activities that are not available to a solitary child.

5. Conclusion

The contrasting semantic preferences of *child* vs. *children* as revealed in this study are unexpected in light of the traditional discussions of the singularity and plurality of count nouns: psychological attributes vs. non-psychological attributes; expressive, empathetic stance vs. expressively neutral stance; alienable vs. inalienable possession; motion vs. stationary predicates. Where possible, I have tried to relate the findings from the collostructional analysis to broader cognitive and behavioural realities such as the nature of empathy and individuals, the role of THE HUMAN BODY and FAMILY cognitive models, the role of 'running around' in the lives of children, etc.

The semantic contrasts that have been revealed in this study involve differences at a constructional level accompanying the use of *child* vs. *children*, differences that go beyond our normal understanding of *children* simply as a plural (= 'more than one') of *child*. A reviewer posed the question whether one can speak of meaning-shift in forming the plural *children*, analogous to how the plural count noun *wines* is said to involve a meaning-shift from the singular mass noun *wine* (Chichiera, 1998; Rothstein, 2010; Katz and Zamparelli, 2012). In my view the difference should be understood as a difference in the potential elaboration of each of the two words at the constructional level, rather than meaning-shift as such. This contrasting behaviour can be further understood by reference to Langacker's (1991: 74-81) discussion of 'types' of number. Langacker's discussion is of relevance in so far as it provides some theoretical scaffolding that allows for the possibility of distinct behaviours associated with a singular vs. plural difference (beyond predictable grammatical rules associated with number in a language). Langacker's argumentation and analysis of nouns is subtle in its detail and a full account of his ideas lies outside the scope of the present study. However, one basic idea underlying that discussion is that singular and plural categories of nouns (including count nouns) are to be strongly differentiated in terms of the 'types' they represent: "[…] it is essential to realize that a singular noun and its corresponding plural represent distinct categories and that what counts as an instance is consequently very different in the two cases" (p. 78). In an example taken from Langacker (1991: 81), *the seven pebbles* is said to profile one instance of the *pebbles* type, not seven instances of *pebble*. *Pebble* and *pebbles* each has its own

distinct type – a 'discrete entity' type for the singular vs. a 'replicate mass' type for the plural. Langacker's approach is a framework for description rather than a predictive theory and does not predict any particular constructional differences between singular and plural types. Nevertheless, it does provide a starting point for appreciating how *child* and *children* can (but not necessarily must) co-occur with such different collexemes and such different semantic preferences in constructions. Singular number represents a 'discrete entity' type with its own preferred constructional properties (preference for prototypically unique referents, greater potential for expression of empathy, etc.). Plural number is associated with its own 'replicate mass' type that has less tolerance for prototypically unique referents, is less likely to be a locus for the expression of empathy and is more likely to refer to emotionally less loaded categories of description, etc.

Obviously, a single study of just *child* vs. *children* in one corpus of English can only take us so far and the results are not immediately generalizable. Some of the patterns discovered here are closely connected with cognitive models of FAMILY and THE HUMAN BODY and are unlikely to find their way into patterning with lexical items outside of those domains. The value in working with one single pair of forms lies in discovering micro-detail that typically goes unnoticed, suggesting the potential value of carrying out other such studies, especially within the lexical domain of kinship terms.

Finally, I hope to have shown that singular vs. plural inflectional differences in nouns warrant closer study than they usually receive. Studying lexicon at the level of lemma, as opposed to inflected word forms each in their own right, may indeed be appropriate for some purposes. As noted in Section 1, Gries (2011) concluded that for the most part, a lemma-based collostructional analysis was sufficiently informative in his investigation of a specific syntactic alternation of English. Notwithstanding Gries' findings related to argument structure of verbs, there is still a danger that describing collostructional behaviour at the lemma level may obscure patterns of interest at the inflected word level of other construction types, just as relying on a mixture of distinct genres may obscure genre-specific patterns.

References

Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Bradford, Barbara Taylor. 1996. *Her Own Rules*. New York: HarperCollins.

Brandon, Jay. 1999. *Angel of Death*. New York: Forge.

Chierchia, Gennaro. 1998. Plurality of mass nouns and the notion of semantic parameter. In Susan Rothstein (ed.), *Events and Grammar*, 53-103. Dordrecht: Kluwer Academic.

Corbett, Greville. 2000. *Number.* Cambridge: Cambridge University Press.

Corsaro, William. 1986. Routines in peer culture. In Jenny Cook-Gumperz, William A. Corsaro, and Jürgen Streek (eds.), *Children's Worlds and Children's Language*, 231-252. Berlin, New York and Amsterdam: Mouton de Gruyter.

Cuff, Benjamin, Sarah J. Brown, Laura Taylor, and Douglas Howat. 2016. Empathy: A review of the concept. *Emotion Review* 8: 144–153.

Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Available online at https://www.english-corpora.org/coca/.

Davies, Mark and Dee Gardner. 2010. *A Frequency Dictionary of Contemporary American English: Word sketches, Collocates, and Thematic Lists.* London and New York: Routledge.

Deignan, Alice. 2006. The grammar of linguistic metaphors. In Anatol Stefanowitsch and Stefan Th. Gries (eds.), *Corpus-based Approaches to Metaphor and Metonymy*, 106-122. Berlin & New York: Mouton de Gruyter.

Flach, Susanne. 2017. *Collostructions: An R implementation for the family of collostructional methods*. Package version v.0.1.0, https://sfla.ch/collostructions/.

Gallager, Shaun. 2005. *How the Body Shapes the Mind*. Oxford and New York: Oxford University Press.

Gries, Stefan Th. 2011. Corpus data in usage-based linguistics: What's the right degree of granularity for the analysis of argument structure constructions? In Mario Brdar, Stefan Th. Gries, and Milena Žic Fuchs (eds.), *Cognitive Linguistics: Convergence and Expansion*, 237-256. Amsterdam and Philadelphia: John Benjamins.

Gries, Stefan Th. 2012. Corpus linguistics: Quantitative methods. In Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*, 1380-1385. Oxford: Wiley-Blackwell.

Gries. Stefan Th. 2019. 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics* 24.3: 385-412.

Gries, Stefan Th. and Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics* 9.1: 97-129.

Hieber, Leanna Renee. 2016. *Eterna and Omega*. New York: Tom Doherty Associates (A Tor Book).

Hillman, Pam. 2011. *Stealing Jake.* Tyndale House ebook.

Hoey, Michael. 1991. *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Hoey, Michael. 2008. Corpus linguistics and word meaning. In Anke Lüdeling and Merja Kytö (eds.), *Corpus Linguistics: An International Handbook*, 972-987. Berlin and New York: Walter de Gruyter.

Hoffman, Alice. 2009. *The Story Sisters.* New York : Shaye Areheart Books.

Holmes, Gina. 2010. *Crossing Oceans.* Carol Stream, IL: Tyndale House Publishers.

Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Katz, Graham and Roberto Zamparelli. 2012. Quantifying count/mass elasticity. In Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz, and Alex Trueman (eds.), Proceedings of the 29th West Coast Conference on Formal Linguistics, 371-379. Somerville, MA: Cascadilla Proceedings Project.

Katz, Graham and Roberto Zamparelli. ms. Meaning-shifting plurality and the count/mass distinction. Available at:

https://edoc.hu-berlin.de/bitstream/handle/18452/2026/katz.pdf?sequence=1

Lakoff, George. 1987. *Women, Fire and Dangerous Things: What Categories Reveal About the Mind.* Chicago: The University of Chicago Press.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Vol. I: Theoretical Prerequisites.* Stanford, CA.: Stanford University Press.

Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar. Vol. II: Descriptive Application.* Stanford, CA.: Stanford University Press.

Main, Alexandra, Eric A. Walle, Carmen Kho, and Jodi Halpern. 2017. The interpersonal functions of empathy: A relational perspective. *Emotion Review* 9.4: 358-366.

Newman, John. 2010. Balancing acts: Empirical pursuits in cognitive linguistics. In Dylan Glynn and Kerstin Fischer (eds.), *Quantitative Methods in Cognitive Semantics*, 79-100. Berlin and New York: Mouton de Gruyter.

Newman, John and Sally Rice. 2004. Patterns of usage for English *sit*, *stand*, and *lie*: A cognitively-inspired exploration in corpus linguistics. *Cognitive Linguistics* 15.3: 351-396.

Nouwen, Rick. 2016. Plurality. In Maria Aloni and Paul Dekker (eds.), *The Cambridge Handbook of Formal Semantics*, 267-284. Cambridge: Cambridge University Press.

Panther, Klaus-Uwe and Günter Radden (eds.) 2011. *Motivation in Grammar and the Lexicon*. Amsterdam: John Benjamins.

Partington, Alan. 1996. *Patterns and Meanings*. Amsterdam: John Benjamins.

Partington, Alan. 2004. Utterly content in each other's company: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics* 9.1: 131-156.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Radden, Günter and Klaus-Uwe Panther (eds.). 2004. *Studies in Linguistic Motivation*. Berlin and New York: Mouton de Gruyter.

Raymond, Linda. 1994. *Rocking the Babies*. New York: Viking.

Ripley, Alexandra. 1991. *Scarlett*. Time Warner: New York.

Rothstein, Susan. 2010. Counting and the mass-count distinction. *Journal of Semantics* 27.3: 343–397.

Shonk, Katherine. 2010. *Happy Now?* New York: Farrar, Straus and Giroux.

Sher, Ira. 2003. *Gentlemen of Space*. New York: Free Press.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Smith, Charlie. 1990. *The Lives of the Dead*. New York: Linden Press/Simon and Schuster.

Stefanowitsch, Anatol and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8.2: 209-243.

Stubbs, Michael. 1995. Collocations and semantic profiles. *Functions of Language* 2: 23–55.

Stubbs, Michael. 2001. *Words and Phrases*. Oxford: Blackwell.

Stubbs, Michael. 2016. Corpus Semantics. In Nick Riemer (ed.), *The Routledge Handbook of Semantics*, 106-121. London and New York: Routledge.

Talmy, Leonard. 2003. *Toward a Cognitive Semantics. Vol. 1: Conceptual Structuring Systems*. Cambridge, Mass. and London, England: The MIT Press.

Thomas, Debbie Fuller. 2008. *Tuesday Night at the Blue Moon*. Chicago: Moody Publishers.

Tiersma, Peter Meijes. 1982. Local and general markedness. *Language* 58.4: 832-849.

White, Susan Rebecca. 2009. *Bound South.* New York: Simon and Schuster.

Appendix 1. Adjectives significantly attracted to the Adjective + *child* construction

	COLLEX	O with *child*	E with *child*	O with *children*	E with *children*	ASSOC	COLL. STR	SIGNIF	Shared
1	*only*	347	220.8	33	159.2	*child*	230.16	*****	Y
2	*middle*	22	13.9	2	10.1	*child*	13.68	***	Y
3	*petulant*	11	6.4	0	4.6	*child*	11.97	***	Y
4	*dear*	20	12.8	2	9.2	*child*	11.85	***	N
5	*sleeping*	21	14.5	4	10.5	*child*	7.83	**	Y
6	*single*	7	4.1	0	2.9	*child*	7.61	**	Y
7	*oldest*	18	12.2	3	8.8	*child*	7.58	**	Y
8	*youngest*	34	25.6	10	18.4	*child*	7.24	**	N
9	*frightened*	20	13.9	4	10.1	*child*	7.09	**	Y
10	*poor*	37	28.5	12	20.5	*child*	6.59	*	Y
11	*holy*	6	3.5	0	2.5	*child*	6.52	*	N
12	*injured*	6	3.5	0	2.5	*child*	6.52	*	N
13	*sensitive*	6	3.5	0	2.5	*child*	6.52	*	N
14	*precious*	13	8.7	2	6.3	*child*	5.84	*	Y
15	*willful*	5	2.9	0	2.1	*child*	5.43	*	N
16	*wretched*	5	2.9	0	2.1	*child*	5.43	*	N
17	*favorite*	11	7.6	2	5.4	*child*	4.28	*	Y
18	*pretty*	8	5.2	1	3.8	*child*	4.16	*	Y
19	*spoiled*	8	5.2	1	3.8	*child*	4.16	*	Y
20	*eldest*	13	9.3	3	6.7	*child*	3.91	*	Y

Appendix 2. Adjectives significantly attracted to the Adjective + *children* construction

	COLLEX	O with *child*	E with *child*	O with *children*	E with *children*	ASSOC	COLL. STR	SIGNIF	Shared
1	*(an)other*	124	203.9	227	147.1	*children*	81.86	*****	Y
2	*older*	8	27.9	40	20.1	*children*	35.49	*****	Y
3	*little*	19	43.6	56	31.4	*children*	33.88	*****	Y
4	*younger*	15	35.4	46	25.6	*children*	28.76	*****	Y
5	*young*	38	61.6	68	44.4	*children*	21.90	*****	Y
6	*local*	0	4.6	8	3.4	*children*	13.94	***	N
7	*sticky*	136	160.4	140	115.6	*children*	9.45	**	N
8	*school-age*	2	7	10	5	*children*	8.79	**	N
9	*remaining*	0	2.9	5	2.1	*children*	8.71	**	Y
10	*small*	0	2.9	5	2.1	*children*	8.71	**	Y
11	*future*	2	5.8	8	4.2	*children*	6.10	*	Y
12	*murdered*	1	4.1	6	2.9	*children*	5.79	*	Y
13	*Jewish*	1	4.1	6	2.9	*children*	5.79	*	Y
14	*white*	2	5.2	7	3.8	*children*	4.83	*	Y
15	*innocent*	12	18	19	13	*children*	4.75	*	Y
16	*deaf*	1	3.5	5	2.5	*children*	4.39	*	Y

Appendix 3. Nouns significantly attracted to the *child's* + N construction

	COLLEX	O with *child's*	E with *child's*	O with *child-ren's*	E with *child-ren's*	ASSOC	COLL. STR.	SIGNIF	Shared
1	*life*	22	11.8	0	10.2	*child's*	27.66	*****	N
2	*face*	21	11.3	0	9.7	*child's*	26.39	*****	N
3	*play*	20	10.8	0	9.2	*child's*	25.11	*****	N
4	*hand*	15	8.1	0	6.9	*child's*	18.77	****	N
5	*voice*	15	8.1	0	6.9	*child's*	18.77	****	N
6	*name*	12	6.5	0	5.5	*child's*	14.99	***	N
7	*body*	11	5.9	0	5.1	*child's*	13.73	***	N
8	*death*	10	5.4	0	4.6	*child's*	12.48	***	N
9	*head*	10	5.4	0	4.6	*child's*	12.48	***	N
10	*game*	9	4.8	0	4.2	*child's*	11.22	***	N
11	*mother*	22	14.5	5	12.5	*child's*	9.33	**	Y
12	*forehead*	7	3.8	0	3.2	*child's*	8.72	**	N
13	*bed*	6	3.2	0	2.8	*child's*	7.47	**	N
14	*drawing*	6	3.2	0	2.8	*child's*	7.47	**	N
15	*hair*	10	5.9	1	5.1	*child's*	7.30	**	Y
16	*toy*	9	5.4	1	4.6	*child's*	6.25	*	Y
17	*cry*	5	2.7	0	2.3	*child's*	6.22	*	N
18	*eyes*	13	8.6	3	7.4	*child's*	5.38	*	Y
19	*father*	9	5.9	2	5.1	*child's*	3.85	*	Y

Appendix 4. Nouns significantly attracted to the *children's* + N construction.

	COLLEX	O with *child's*	E with *child's*	O with *child-ren's*	E with *child-ren's*	ASSOC	COLL. STR.	SIGNIF	Shared
1	*books*	0	15.1	28	12.9	*children's*	44.11	*****	N
2	*hospital*	0	11.8	22	10.2	*children's*	34.51	*****	N
3	*book*	0	9.7	18	8.3	*children's*	28.15	*****	N
4	*voices*	0	7	13	6	*children's*	20.26	*****	N
5	*faces*	0	4.8	9	4.2	*children's*	13.99	***	N
6	*names*	0	4.8	9	4.2	*children's*	13.99	***	N
7	*children*	0	3.8	7	3.2	*children's*	10.86	***	N
8	*lives*	0	3.2	6	2.8	*children's*	9.31	**	N
9	*toys*	1	5.4	9	4.6	*children's*	8.71	**	Y
10	*clinic*	0	2.7	5	2.3	*children's*	7.75	**	N
11	*section*	0	2.7	5	2.3	*children's*	7.75	**	N
12	*ward*	0	2.7	5	2.3	*children's*	7.75	**	N
13	*clothes*	1	4.3	7	3.7	*children's*	6.06	*	Y

Appendix 5. Verbs attracted to the *child* + Ving vs. *children* + Ving constructions

	COLLEX	O with *child*	E with *child*	O with *children*	E with *children*	ASSOC	COLL. STR	SIGNIF	Shared
1	*trying*	4	1.8	1	3.2	*child*	3.93	*	Y
2	*working*	4	1.8	1	3.2	*child*	3.93	*	Y
3	*growing*	7	4.1	4	6.9	*child*	3.29	ns	Y
4	*lying*	3	1.8	2	3.2	*child*	1.10	ns	Y
5	*clinging*	3	2.2	3	3.8	*child*	0.43	ns	Y
6	*standing*	7	6.3	10	10.7	*child*	0.13	ns	Y
7	*holding*	3	2.6	4	4.4	*child*	0.10	ns	Y
8	*going*	2	1.8	3	3.2	*child*	0.02	ns	Y
9	*crying*	3	3	5	5	*child*	0.00	ns	Y
10	*singing*	2	2.2	4	3.8	*children*	0.04	ns	Y
11	*sitting*	5	6.3	12	10.7	*children*	0.45	ns	Y
12	*looking*	1	1.8	4	3.2	*children*	0.69	ns	Y
13	*playing*	13	17	33	29	*children*	1.74	ns	Y
14	*living*	2	4.8	11	8.2	*children*	3.05	ns	Y
15	*screaming*	0	2.2	6	3.8	*children*	5.59	*	N
16	*coming*	0	2.6	7	4.4	*children*	6.53	*	N
17	*running*	2	8.5	21	14.5	*children*	10.22	**	Y