Constructing Age in Children's Literature: A Digital Approach to Guus Kuijer's Oeuvre

Wouter Haverals and Vanessa Joosen

In recent years, digital humanities (DH) has left a clear mark on the study of literature. The aid of computer algorithms facilitates the analysis of collections of literature that are so vast that they transcend the limits of what a person could possibly read during a lifetime. Moreover, computational power offers the prospect of tackling new and exciting research questions, which were previously bound to the storage capacity and working memory of the human researcher. Because of the limits to human performance, a quantitative approach—whereby literature is condensed into "data to be examined"—can help to break new ground. Obviously, engaging in digital humanities research requires literature to be available in digital, machine-readable form (Owens; Jockers 7–8). While initiatives to digitize children's books have led to impressive online collections, such as the Baldwin Library of Historical Children's Literature, the Auslit digital collection of storytelling, and the Digital Library for Dutch Literature (DBNL),[1] most children's literature scholars have so far shied away from applying digital techniques in their analyses, with a few notable exceptions (Karsdorp; Fitzsimmons and Alteri; Cross et al.). *The Edinburgh Companion to Children's Literature* (2017) ranked distant reading under its section of "Unmapped Territories" (Giddens) and digital humanities still rarely features in journals and conference papers in this field. Children's literature studies thus risks separating itself from one of the most significant shifts in twenty-first-century literary analysis. In this article, we will highlight some of the potential gains of digital humanities for the analysis of children's literature and address some of the challenges and limits in such computational approaches.

Moving away from a selection of texts to a large body of relevant literature implies a shift from "close" to "distant reading" (Moretti), or from "microanalysis" to "macroanalysis" (Jockers). Macroanalysis reflects the increase of scale with which research in DH is carried out. This paradigm

The Lion and the Unicorn 45 (2021) 25–45 © 2021 by Johns Hopkins University Press

shift is sometimes mistrusted (Da), scorned (Brennan), or even perceived as hostile toward traditional literary scholarship (Fish; Kirsch). In *Distant Horizons* (2019), digital humanist Ted Underwood attempts to soothe these concerns and grievances. He argues that it is extremely challenging for a single scientist to identify and fully understand long-term developments in literature, drawing a comparison with the observation of the Earth's curvature: "A single pair of eyes at ground level can't grasp the curve of the horizon, and arguments by a single reader's memory can't reveal the largest patterns organizing literary history" (ix–x). Computers are better equipped for this task, Underwood convincingly claims, provided that they are operated in a clever way. What is deemed clever use of computer algorithms, and what is not, is—unsurprisingly—a matter of debate. Relying on binary terminology—"close" versus "distant reading," and "micro-" versus "macroanalysis"—has inadvertently prompted the rise of a dichotomy in DH studies (Underwood 2–3). Some researchers, like Franco Moretti (2000), insist that a strict division between the two is inevitable. "[T]he more ambitious the project, the greater must the distance be," Moretti (57) argues. For other researchers, distant and close reading are not mutually exclusive (Rosen; Jänicke et al.; Berglund; van der Ven et al.). They advocate an integrated approach that enriches literary analyses—an approach that we subscribe to in this article.

Macroanalysis has already contributed to, amongst others, the assessment of gender and its relation to word choice in literary texts (Jockers; Jockers and Mimno). In this article, we apply methods from digital humanities to gain more insight into the ideological construction of age in the oeuvre of a single author. This is a relatively small set compared to the 3,279 books that Matthew Jockers and David Mimno (2013) have "macroanalysed," and one could argue that an in-depth study of a single author's oeuvre still falls within the span of what a human mind can master. Nevertheless, we will show that digital tools can facilitate the analysis of trends in the explicit ideology in the oeuvre of a single author. Moreover, by using word calculations and counting combinations of words, they can trace implicit patterns that are impossible to perceive by the human mind, even in a corpus that is relatively small according to DH standards. In this respect, we subscribe to the idea that "macro- or distant-scale perspectives on literature offer scholars a necessary context for and complement to closer readings" (Jockers and Mimno 768).

Our analysis focuses on the Dutch author Guus Kuijer. Kuijer has developed a substantial oeuvre that stretches over various decades. He has earned both national and international recognition for the quality of his work in children's literature, his books have been widely read by children in the Netherlands and several titles have been translated and adapted to other media.[2] Moreover, the concept of "age" is relevant to Kuijer's oeuvre in various ways:

he is a crosswriter who has authored fiction for children, adolescents and adults, and intergenerational relationships are a recurrent thematic feature in his work (Joosen, *Adulthood*). Since discussions on age in his works have so far been limited to case-based research, this article seeks to offer a fuller understanding of the role that age plays in Kuijer's oeuvre. What explicit ideas about age do his books offer? What implicit age norms can be derived from the characterization of the figures in his works? And to what extent does the age category of the intended reader determine the form and themes of Kuijer's fiction?

Explicit Reflections on Age: Method and Analysis

As previously mentioned, we use a combination of distant and close reading strategies to answer the research questions outlined above. To do so, we first either obtained or created digital text files for all thirty-two novels and collections of short stories from Kuijer's oeuvre.[3] We received some novels in digital format through the publisher or DBNL. Others we scanned ourselves and converted with optical character recognition software to computer-readable text. For some aspects of our research we could simply use the plain textual material that results from such digitization, but for most purposes introduced in this article, we converted the plain text to XML, a format that allows users to enrich the text with human-coded annotations.[4] These annotations are inserted in so-called "tags," that function as labels and help users extract specific elements of the text in a later stage. Depending on the information that is put in the tags, researchers can label and automatically extract fragments related to their specific focus. For instance, they can annotate and then extract all sections related to descriptions of landscape, animated objects or animals, or all explicit ideas about gender, or all occurrences of Latinx characters, etc. Once a method for tagging has been agreed, annotation tasks can be divided over a group of researchers, including students and volunteers, so that it is possible to carry out an analysis over a larger corpus without having one single researcher read all the texts.

For the results presented in this article, a team of researchers "tagged" what Joosen (*Adulthood* 7) calls "metareflections" on age: passages in which an age norm is made explicit or where generalizing thoughts about an age or life stage are expressed. Such passages are enclosed with so-called "<seg>-tags," which are supplemented with information about the age category concerned. The following examples from Guus Kuijer's *Krassen in het tafelblad* (1978, Scratches in the table-leaf) are metareflections on old age, adulthood and childhood respectively:[5]

<seg about="oldadult">Old people die, that's just the way it is.</seg>
<seg about="adult">because grownups, well, you know, they are easily shocked</seg>
<seg about="child">They can be troublesome, can't they, at that age.</seg>

From the annotated books we exported all metareflections and their corresponding information on age to a spreadsheet with the aid of a computer script that we developed ourselves. Additionally, this information was enriched with the title of the book and the age of the intended reader as listed in the Central Dutch Library Catalogue. Filters then allowed us to create subsets for specific categories (e.g., all metareflections for Kuijer's juvenile fiction; all metareflections on old age, etc.).

When considering the table with all the tagged metareflections for Kuijer's oeuvre, one of the first striking observations is the high amount: 353 metareflections for 32 books (which all together account for 978,057 words), or 3.6 for every 10,000 words. For comparison: the *Harry Potter* series also counts roughly one million words (1,084,170 to be precise), but contains only 67 metareflections or 0.6 for every 10,000 words, six times less than Kuijer's books. The distribution of metareflections offers further insights into the role that age plays as an overt theme in Kuijer's oeuvre. We find metareflections on all age categories in Kuijer's children's, adolescent and adult books. But the quantity, quality, and content differ. 259 can be found in his children's and adolescent fiction (27 titles, 635,991 words) and 94 in his books for adults (5 titles, 342,066 words). If we calculate the relative number of metareflections, we see that Kuijer reflects significantly more on age in his juvenile fiction than in his adult works: 4 metareflections for every 10,000 words versus 2.7 respectively.

An overview broken down by the age of the intended reader (juvenile versus adult literature) shows that both in his children's books and his adult work, Kuijer reflects more on youth than on adulthood (Figure 1).

However, when considering their content, we see that the metareflections in Kuijer's adult books often evoke childhood to say something about adulthood, rather than about childhood. For example:

> When someone wants to leave they duck and crawl out **like a child** between the legs. (*De man met de hamer*)[6]
> I cried **like a small child** at Efua's breasts (*De redder van Afrika*)[7]
> Nasta recognized **the child in her** that looks with longing eyes at a birthday gift still wrapped. (*Het vogeltje van Amsterdam*)[8]

Each time, childhood is invoked in a simile or as a point of reference to say something about adulthood. Two-thirds of the metareflections on childhood in Kuijer's adult fiction are of this type; only roughly a third relates to childhood itself. By contrast, none of the metareflections on adolescence

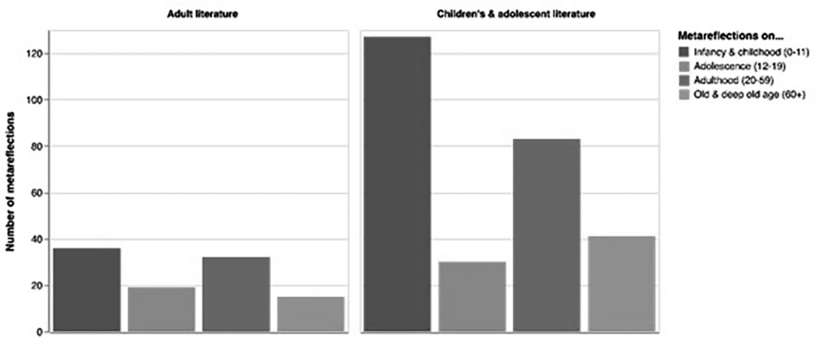

Figure 1. Metareflections on specific age groups in the work of Guus Kuijer, ordered by the age of the intended reader.

are invoked to say something about another stage in life. The survey shows that childhood in Kuijer's adult works acts as a recurrent point of reference to express feelings of vulnerability, powerlessness, and innocence that also occur in other stages in life. Scholars have so far mostly put the emphasis of childhood as a root metaphor that is mapped onto old age (Hockey and James; Joosen, *Connecting*). In Kuijer's adult books, childhood is invoked frequently to express ideas about adulthood, even more so than about old age, which features less as a theme in his adult fiction.

In addition, the overview makes clear that adulthood is a recurrent explicit theme in Kuijer's juvenile fiction. There, most metareflections contain assertions about adulthood in its own right; only a few times is children's behavior measured by the standard of adulthood. Variations of the phrase "You are acting like a grown up!" occur three times in Kuijer's children's books.[9] The only other time when a metareflection about adulthood is used as a simile for a child's behavior it is to describe a kind of walking that the children usually find boring: "We walked like grownups: with no purpose." (*Met de wind mee naar zee*).[10] The overview of metareflections thus shows that Kuijer's children's and adolescent books are more invested in reflecting on other stages in life than his adult literature, which is first and foremost concerned with exploring adulthood itself. Within Kuijer's oeuvre, his juvenile fiction is the prime place to explicitly reflect on the life course and the meaning of age in a broad sense. This observation highlights the importance of including children's literature in age studies, and vice versa.

What about the content of the metareflections? Since there are so many and they are so diverse, we will limit ourselves here to a discussion of the metareflections on adulthood in a narrow sense (excluding old age) as they are expressed in Kuijer's oeuvre. Two-thirds of these express a negative or

restrictive idea about adulthood; only a third offers neutral and positive statements. The negative statements apply to a range of adult features, such as repulsive aspects of the adult body, a lack of understanding, strange habits and silly clothes, but most are related to a lack of control over feelings: the idea that adults are easily shocked, frightened and angered, and that they cannot endure pain. For example:

> Adults, they are so easily shocked, one day they are going to shock themselves to misery![11] (*Krassen in het tafelblad*)
>
> Sometimes adults don't think properly before they say something.[12] (*Met de wind mee naar zee*)
>
> Adults often make fuss about nothing.[13] (*Tin Toeval en de kunst van het verdwalen*)

In addition, Kuijer's fiction expresses various limiting age norms for what adults can or should (not) do: for example, that adults should be married, are incapable of change, lack creativity, must not be afraid, do not like playing in the street and should be responsible.

With all the negative and restrictive age norms that dominate explicit reflection on adulthood, Kuijer's works can be argued to contribute to a general aversion to adulthood that Susan Neiman has thematized in *Why Grow Up?* (2014). Two considerations complicate such a straightforward conclusion though. On the one hand, the digital tools we applied isolate metareflections from their context. When matched with close reading strategies, we see that many of the negative assertions either appear in fights where insults go both ways, or that they are uttered by characters who are marked as unreliable. As Joosen (*Adulthood* 108) has argued for Kuijer's Polleke, for example, the child's disgust of the adult body is so extreme that it serves to mock *her* at least as much as the adults she resents. Nevertheless, such parodies of an aversion to adulthood are reminiscent of what Linda Hutcheon (16–17) has described as the double bind of all parodies: they raise up what they criticize in the very moment of bringing it down. After all, only what is relevant to a certain culture is worth the attention that parody pays to it. Kuijer's metareflections thus affirm the negativity associated with adulthood even as they question it.

Implicit Age Norms in Direct Speech: Method and Analysis

Explicit reflections on ages and life stages are not the only way in which novels convey age norms. In fact, John Stephens argues that explicit ideology is far less effective in socializing readers than implicit ideology (10). While the former may invite debate and criticism, the latter is often naturalized and requires critical reading strategies to be recognized and questioned.

Can computer tools assist researchers in identifying implicit patterns in age norms? Here, we turn to language, which is one way of characterizing literary figures implicitly. The scatterplot analysis that we will present in this section offers a kind of "topic modelling" that can be used to investigate how the direct speech of a group of characters distinguishes itself from others. The method that we describe can help to understand and visualize, among others, how girls' speech differs from boys' speech, or whether characters' race and ethnicity lead to a distinct vocabulary. We will use this method to get a better insight into the construction of an age ideology through language, asking the following questions in particular: does children's speech in Kuijer's books differ from that of adults, and if so: how? We examine more specifically whether juvenile and adult characters use different vocabularies. Can we derive anything from those findings about the topics they like to discuss and the way the stages in life are shaped?

Again, we first resort to digital tools to lay the basis for our analysis. In our annotated XML-files, we make a distinction between direct and indirect speech. For passages told by a narrator in the third person we use the tag <said direct="false">. For dialogue and passages told by a character, we replace "false" with "true" and specify the speaker with a so-called "character ID." This is a lowercase, unspaced label that annotators create themselves to identify a specific character unambiguously. A dialogue introduced by the narrator then looks as follows:

> <said direct="false">Grandpa is staying with Madelief. That's nice, because Madelief has a break and her mother hasn't. With grandpa there it's a lot nicer. For example, you can play games with him very well. When he loses he only gets a little angry.</said>
> <said direct="true" who="madelief">"Grandpa?"</said>
> <said direct="false">Madelief asks.</said>
> <said direct="true" who="madelief">"Did you think grandma was a nice person?"</said>[14]

For each book, we made a spreadsheet that matches the character IDs (e.g., madelief) with a set of identity features, including age,[15] gender, and race/ethnicity. We use exact numbers when a character's age is explicitly mentioned or can be derived, and when that information is lacking, we refer to life stages that are rendered as refined as possible. In addition, we have developed an age model that helps us to standardize the way the annotators attribute age to a character. For the categorisation and names we used models from age studies, such as Lorraine Green's (2010) and Thomas Armstrong's (2019), as reference points (see Figure 2).

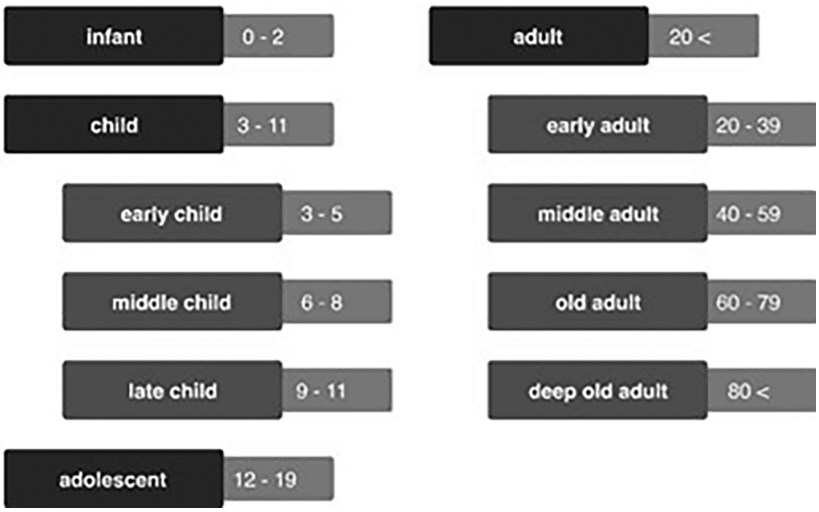

Figure 2. Categorical organization of age groups.

For the attribution of age, we rely on the information that the texts offer. Sometimes a character's age or life stage is mentioned explicitly. At other times, we rely on context: characters attending primary school are categorized as children in the broad sense, parents are labelled as adults and grandparents as old, unless more specific information is given. We realize that this type of labelling does not cover the diversity of age in reality: some people have children before the age of 20 and not all grandparents have reached the age of 60. One of the downsides of the model is that we reinforce the age norms that we want to question to a certain extent, and that we overemphasize the importance of numerical age. Such reductions and artificial separations are needed, however, in order to prepare the material for digital analysis, which typically works with countable entities.

For the analysis of implicit age norms in direct speech, we developed a computer script for the extraction of all the text that we labeled with <said direct="true" who="character ID">-tags. The script matches the character ID in the tags with the identification markers listed in our spreadsheet, in particular the age that was attributed to the character ID. For the rougher analysis of age that we want to carry out here, we reduce the specific age categories in our model by grouping them under two broad classes: child and adolescent on the one hand, and adult and old adult on the other. Then every text that is marked as direct speech is grouped under the right age category via the character ID, so that the script creates one text file containing all the direct speech uttered by child and adolescent characters, and another text file

with all the direct speech of adult and old figures. Next, we apply Scattertext (Kessler), a powerful open-source tool for visualizing linguistic variation in different text sets. One of the main challenges when investigating the differences between large corpora lies in finding a balance between an attractive visualization and the possibilities it offers for interpretation. Scattertext meets this challenge in an inventive manner. The tool is able to render a stand-alone HTML-file, which lends itself for a query-based investigation.

Figure 3 shows the result of our analysis, comparing adult and old adult speech on the one hand, and child and adolescent speech on the other.[16] The higher up a dot is on the y-axis, the more children or adolescent characters use it. Similarly, the further right on the x-axis a dot appears, the more its corresponding word is used by an (old) adult character. Highly frequent words in both age categories appear in the upper right-hand corner. The top-left zone contains words that are often used by children, but never or hardly ever by adults; whereas the bottom-right zone contains words that feature regularly in adult speech, yet are rare or absent in child speech. By mousing-over the individual dots, statistics can be retrieved about the word's relative occurrence in the two distinct categories. Furthermore, some context is also presented for each word, as well the characters who say it. This way, a more in-depth analysis can be pursued into the specificities of the vocabulary used by both age groups.

When analyzing the results, it is not just interesting to see what divides child and adult speech, but also what they have in common. In the shared zone, the word "ruzie" (fight, argument) features high. Indeed, Kuijer started writing in the 1970s when intergenerational conflict was a common theme in Dutch children's literature. Characters in his early work often have strong arguments, and this trend has continued also in his later work. But from the children's and adult speech, we can derive that the two age groups do not argue in quite the same way. What distinguishes the children's speech is intensity and colloquialism: intensifiers like "zooo" (sooo) and "hartstikke" (totally), strong deniers like "nietes" (no way), as well as derogatory terms like "slome" (lame), "achterlijk" (retarded), "schijterd" (you chicken), "stom / stomme" (stupid) and "rolmops" (rollmop) are used by various child characters and point at children's anger and frustration when they are in disagreement. Although the words "trut" (cow) and "zak" (jerk) occur in both child and adult speech, adults also tend to use a more persuasive discourse of reasoning through the words "nietwaar" (isn't it), "bovendien" (moreover), "absoluut" (absolutely), and "vertrouwen" (trust), as well the interjection "tja" (well, h'm), which can both signal agreement or a diplomatic way of expressing disagreement and acceptance. This forms an interesting contrast with the metareflections discussed before, where adults are accused of being unable to

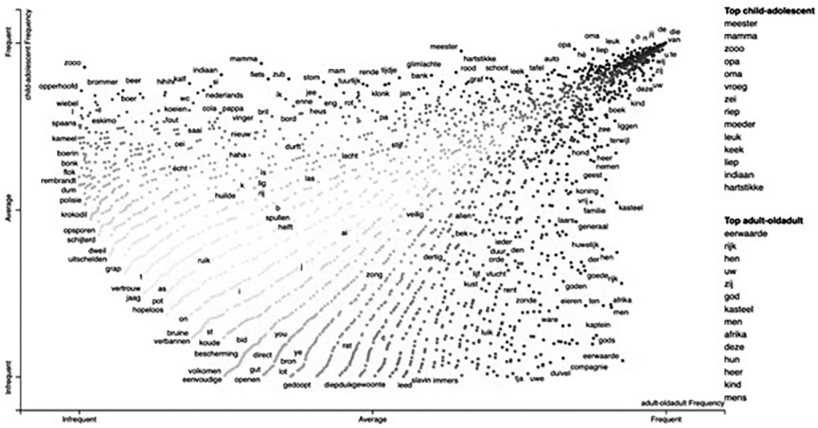

Figure 3. Scattertext visualization of direct speech in novels by Guus Kuijer. The analysis offers a comparison between child and adolescent speech on the one hand, and adult and old adult speech on the other.

control their feelings. Their manner of conversation in the corpus as a whole suggests the contrary. An additional striking observation is the occurrence of the word "kinderachtig" (childish). Somewhat surprisingly, this word is far more used by children than by adults, where this term is rare. It suggests that in Kuijer's works, children guard and reinforce age norms about childhood more strongly than adults. Finally, if we look at the context and speakers of some words, we see that age intersects with other identity markers when it comes to vocabulary. For example, the scatterplot shows an interesting difference in onomatopoeias used for laughter: "hihihi" has a high frequency among children only, while "hahaha" is more equally distributed among children and adults. The results are at the same time clearly gendered: most of the children giggling "hihihi" are girls, with only a few boys using it, while "hahaha" occurs across gender and age. More sophisticated statistical analysis on a bigger data set can help us trace patterns through various identity categories and pick up on intersectionality on more levels. If enough annotated textual material is available, a scatterplot would allow, for example, to contrast the speech of more specific groups, e.g., twelve-year-old female figures versus male characters in the same age range, or the speech of black characters developed by black authors versus those developed by white authors.

Stylometric Analysis: Method and Analysis

With the analysis first of metareflections and then implicit age norms through speech, we are gradually zooming out. Using techniques from the field of computational stylometry, it is also possible to investigate Kuijer's oeuvre as a whole. Broadly speaking, this field of research engages in the quantitative analysis of stylistic features. In this sense, computational stylometry is an attempt at the statistically underpinned "measurement" of authorial style (Holmes; Daelemans). In literary studies, its best known applications have come from attributing anonymous and pseudonymous texts to their authors (Kestemont). For example, stylometry provided convincing evidence to the claim that J. K. Rowling was the author behind Robert Galbraith (Juola). In this section, we investigate what this technique can offer to get a better understanding of how age functions in an author's oeuvre.

For our stylometric analysis of Kuijer's oeuvre, we built a bag-of-words model (BoW), which represents all the words in a given text as vectors, compressing information about their relative frequencies as numerical values (Zhang et al.).[17] Any information about the order of individual words in the novels is discarded (hence the name *bag*-of-words). When building a BoW-model, raw textual material is used. In our case that comprises all of Kuijer's thirty-two novels without annotations or any other scholarly interference on a textual level. After obtaining the BoW-model, we apply a hierarchical clustering algorithm that produces a cluster tree (also called a *dendrogram*), shown in figure 4.[18] This tree is based on the two thousand most frequent words in Kuijer's entire oeuvre and their relative frequencies in individual titles.[19] We did, however, ensure that little to no weight was given to character names, which might distort the results.[20] In the resulting tree structure, books are grouped closer together if their word uses and corresponding frequencies are similar. The tree thus gives an impression of stylistic and topical resemblance between books. The research questions that underlie our construction of the hierarchical cluster tree are the following: can a stylometric analysis pick up on similarities between books that we know belong together in Kuijer's oeuvre, for instance because they are part of a series? Is the division in children's, adolescent, and adult fiction in Kuijer's oeuvre also apparent from a stylometric analysis? Or is the time period a more determining factor than the age categories of the intended reader?

First of all, in figure 4 we see that books that belong to the same series are indeed grouped together. The titles in the series around Tin, Madelief, and Polleke are all clustered. What is even more striking is the position of *Tin Toeval en de kunst van Madelief* (Tin Coincidence and the art of Madelief), a small booklet that combines the protagonists of two of Kuijer's popular

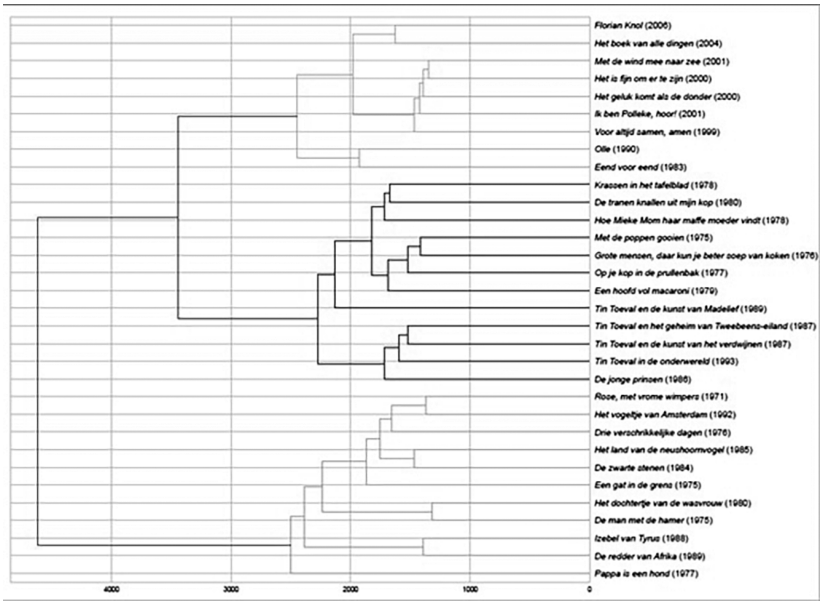

Figure 4. Hierarchical cluster tree, illustrating the similarity between word frequencies in the oeuvre of Guus Kuijer (number of selected features = 2,000; minimum document frequency = 5.).

series. In the tree it appears exactly between the *Madelief* and the *Tin Toeval* books. The *Madelief* group is clustered with two additional titles, *De tranen knallen uit mijn kop* (The tears are bursting out of my head) and *Hoe Mieke Mom haar maffe moeder vindt* (How Mieke Mom finds her mad mother). They are not part of the series but were children's books published in the same period, 1975–80. Similarly, the *Tin* cluster contains one book that is not part of the series, *De jonge prinsen* (The young princes), but that is published a year before the first *Tin* book for the same age group (child readers). A second consistent cluster consists of Kuijer's books from 1999 to 2006 with the *Polleke* books, *Het boek van alle dingen* (The Book of Everything) and *Florian Knol*. In all these cases, ages and periods coincide and this is apparently reflected in a recognizable style and topicality.

The other groups are more diverse and raise interesting questions for further exploration. One children's book, *Eend voor eend* (Duck for duck, 1983), seems somewhat oddly placed with titles from the 1990s and 2000s. Yet, the fact that it is clustered with *Olle* (1990) makes sense. Both are the only titles in Kuijer's oeuvre told by a first-person adult narrator who presents

himself as an alter ego of the author himself. Moreover, both books describe anthropomorphized animals and their behavior in nature. Their topical and narrative similarity is apparently strong enough to supersede temporal differences in Kuijer's use of vocabulary.

The hierarchical cluster tree also shows that there is a clear distinction in Kuijer's oeuvre between his children's books on the one hand, and his fiction for adults and adolescents on the other. As it is apparent from the formation of clusters, his adolescent fiction is closer to his adult fiction than to his books for younger readers, and this division supersedes periods and genres. There is only one exception to this neat division: *Pappa is een hond* (Daddy is a dog) is the only children's book grouped with Kuijer's adolescent and adult fiction. Its unusual position is reminiscent of a remark by Steven Campe. After Kuijer had become a beloved children's author with the *Madelief* books by the late 1970s, some critics argued that he was alienating readers with his more recent books. *Pappa is een hond* is a far stretch from the domestic realism presented in the *Madelief* series. Campe argued that it was really a book for adults disguised as children's literature: "The question is whether you help a child by offering such a story. We immediately believe in Guus Kuijer's integrity, but the question is not answered with that observation. How does the young reader benefit from this dream in daily life?" (cited in Tromp 92–93).[21] Stylometrically speaking, Campe is right in aligning this specific title with Kuijer's work for adolescents and adults.

The cluster suggests that Kuijer's adolescent fiction in general resembles his adult fiction more closely than most of his children's fiction. The organization is not completely random, but is at least partly determined by the age of the intended reader, the period and the genre. Hence, we see a logical combination of *Izebel van Tyrus* and *De redder van Afrika* (Africa's savior), two historical novels for adults from the late 1980s, and of *Het land van de neushoornvogel* (The land of the great hornbill) and *De zwarte stenen* (The black rocks), two books for adolescents that fall within the fantasy genre and were published in the mid-1980s. In all the dendrograms we produced with different variables, *Het vogeltje van Amsterdam* (The little bird of Amsterdam), a novel for adults from 1992, was clustered unexpectedly. In figure 4 it is matched with Kuijer's debut, *Rose, met vrome wimpers* (Pink, with pious eyelashes) and a novel for adolescents for the same decade, *Drie verschrikkelijke dagen* (Three terrible days). In similar analyses with different variables, it was also clustered with adolescent fiction from the 1970s and 1980s. This raises the hypothesis that *Het vogeltje van Amsterdam* may have been written long before it was published. Previous research in stylochronometry has made a convincing case for reconsidering periodization in the oeuvre of, for example, Samuel Beckett (Kestemont and Van Hulle).

Alternatively, one can also speculate whether the unexpectedly clustered novel shares specific thematic and stylistic characteristics with Kuijer's adolescent fiction. We see the stylometric analysis as an invitation to read *Het vogeltje van Amsterdam* in a new intertextual light that prioritizes Kuijer's adolescent fiction as potential pre-texts.

Conclusion

In this article we hope to have demonstrated that the application of techniques from the field of digital humanities can enrich the study of age in children's literature. Part of our endeavor has focused on the interplay between a close reading and a distant reading perspective. In our view, quantitative analyses, involving techniques from the field of DH, do not imply an exclusion, nor a replacement of longer established methods of narrative analysis. Rather, we want to highlight the enhancing and complementary effect of an exchange of insights between the two. Of course, DH has its own challenges and limits. We pointed out that the tools we developed and applied onto the corpus of Guus Kuijer's novels isolate words from their context. Doing so, our analysis of metareflections and direct speech essentially relies on countable instances, which involves the categorization of a reality that is more diverse and messy than any numerical approach can capture. Furthermore, some level of human interpretation is still involved, even when resorting to distant reading, for example to determine the age of a character or to recognize a metareflection. When reading tasks are divided among a group of researchers, as we have done for the analyses presented in this article, minor differences in interpretation may arise and somewhat skew results.

Nevertheless, digital tools can help to verify results from case-based research and allow for an exploration of research questions in a larger corpus. They can also reveal trends that have stayed under the radar or thus far remained elusive from a close reading approach. For example, our research has shown that Kuijer's juvenile literature is the prime place where he reflects on age, which supports Joosen's plea to give children's literature a more central place in age studies, and vice versa, to draw on age studies to study Kuijer's works. Moreover, we see that the negative and restrictive discourse about adulthood that has previously been addressed in selected titles, stretches out over his entire oeuvre. Both the analysis of implicit age norms in the vocabulary that the characters use as a consideration of those negative statements in context serve to put that negativity into perspective, however. Results that we did not expect include the prominence of reflections on childhood in Kuijer's adult work and the fact that childhood is mostly invoked there to express sentiments about adult characters. In this respect, the abundance of

the word "childish" in child speech also came as a surprise. Previously, when childism has been investigated in Kuijer's works, the focus was put on adult characters. A more systematic exploration shows that childist age norms are also phrased by young characters, even if, just like the negative statements about adulthood, they are often peppered with a good deal of benevolent irony. Not only the trends but also the exceptions are interesting, and some of our results raise new research questions about books that we did not realize were exceptional, such as *Het vogeltje van Amsterdam*.

Finally, digital humanities is a field that is rapidly evolving and holds promises for the immediate future of children's literature research, for example in the mining of historical children's literature, where vast amounts of digitized texts are readily available and eagerly awaiting scholarly attention, and in the identification and analysis of visual material, where digital scholars (Ueno) are currently breaking new ground. Stylometry offers further possibilities for identifying anonymous children's texts—of which there are many—or to assess the extent to which children's authors' styles change in the course of their career or when they address specific audiences. These are opportunities not to be missed to link the study of children's literature to the digital shift in the humanities and the innovative horizons this joint venture has to offer.

Wouter Haverals is a Postdoctoral Researcher, working at the Institute for the Study of Literature in the Low Countries (ISNL), and the Antwerp Centre for Digital Humanities and Literary Criticism (ACDC). In his research, he unites two disciplines: traditional philology and computational literary studies. For his PhD thesis, he investigated how artificial intelligence can contribute to an objective reconstruction of the pronunciation of medieval Dutch poetry. Within the project Constructing Age for Young Readers (CAFYR) he focused on applying techniques from the field of digital humanities to reveal the implicit and explicit ideas about age in a large corpus of literary works. Currently, he is working within the project Silent Voices, aimed at uncovering the stylistic features of an extensive corpus of medieval texts.

Vanessa Joosen is Associate Professor of English literature and children's literature at the University of Antwerp. There she leads the ERC-funded project Constructing Age for Young Readers and organizes the annual Children's Literature Summer School. Vanessa Joosen is the author of, amongst authors, Critical and Creative Perspectives on Fairy Tales (Wayne State UP 2011) and Adulthood in Children's Literature (Bloomsbury 2018). She also edited Connecting Childhood and Old Age in Popular Media (U of Mississippi P 2018). Joosen is interested in matching perspectives from age studies with children's literature studies, and in combining digital tools and narrative

analyses to study children's books. In a recent GOA project, she and her colleagues at the University of Antwerp use machine learning to study bias in illustrations to historical children's books.

Acknowledgments

The authors wrote this article as part of the research project "Constructing Age for Young Readers." This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 804920). We want to thank Mike Kestemont and Lindsey Geybels for their technical support, Guus Kuijer for his permission to analyze the digitized works, Querido and DBNL for providing some of the digital texts, and all interns and students who helped digitizing and annotating the books.

Notes

Some additional digital text collections that also offer a smaller selection of digitized children's books are Project Gutenberg, HathiTrust, and Internet Archive.

[2] Awards that Kuijer has received include various Golden Slates (1976, 1979, 2000, 2005) and the German Children's Literature award (1982, 2002) for individual titles, as well prizes for his entire oeuvre, such as the Dutch national award for children's literature (1979), and the Astrid Lindgren Memorial Award (2012). Kuijer won the award of the Dutch children's jury for *Tin Toeval en de kunst van het verdwalen* (de Vries 108). Adaptations include the television series *Madelief* (1993–96), the films *Madelief: Krassen in het Tafelblad* (1998) and *Polleke* (2003) as well as various dramatized performances of *The Book of Everything*.

[3] Our team includes a children's literature specialist, a specialist in digital humanities, two PhD students and a group of students who contributed to the digitization and annotation of the books.

[4] We follow the conventions of the Text Encoding Initiative (TEI) that has developed standardized rules for XML-encoding (see www. https://tei-c.org; TEI Consortium 2019). We supplemented these with rules specific to our project and built a customized schema for this research project.

[5] The </seg> makes clear where the metareflection ends. For the analysis we used the Dutch originals. All translations from Dutch in this article are our own. Original text: <seg about="oldadult">Ouwe mensen gaan dood, zo ís 't nu eenmaal.</seg> <seg about="adult">want grote mensen, nou ja, je weet wel, die schrikken nogal gauw.</seg> <seg about="child">Ze zijn wel es lastig hè, op die leeftijd.</seg>.

[6] "Wanneer iemand weg wil maakt hij zich klein en kruipt als een kind tussen de benen door naar buiten."

[7] "Ik heb als een klein kind gesnikt aan Efua 's borsten."

[8] "Nasta herkende het kind in haar dat begerig naar een nog ingepakt verjaarscadeau kijkt."

[9] "Je lijkt wel een groot mens" in *Tin Toeval en de kunst van Madelief* en *Een hoofd vol macaroni*. "Jullie lijken wel grote mensen!" in *Met de wind mee naar zee* (2001): "jullie lijken wel grote mensen! Bah!"

[10] "We wandelden als grote mensen: nergens heen."

[11] "Grote mensen, die schrikken zo gauw, die schrikken zich nog eens ongelukkig!"

[12] "Soms denken grote mensen niet goed na voordat ze iets zeggen."

[13] "Grote mensen maken zich vaak druk om niks."

[14] <said direct="false">Opa logeert bij Madelief. Dat is fijn, want Madelief heeft vakantie en haar moeder niet. Met opa erbij is het een stuk gezelliger. Je kunt bijvoorbeeld heel goed spelletjes met hem doen. Als hij verliest wordt hij maar een klein beetje kwaad.</said> <said direct="true" who="madelief">"Opa?"</said><said direct="false"> vraagt Madelief.</said> <said direct="true" who="madelief">"Vond jij oma lief?"</said>.

[15] For stories containing flashbacks or stretching out over several years, we create separate IDs for the characters who appear at different ages (e.g., grandpa and grandpayoung). That way we can trace the age for each character in detail.

[16] In order to perform this analysis, highly frequent character names were filtered out and subsequently removed from the direct speech corpora. The occurrence of these names is often limited to a specific novel or a series (e.g., Kuijer's *Tin Toeval* series), and therefore is not indicative of the linguistic variation of *all* speech instances. A total of eighty-two character names were manually removed.

[17] The code used for the stylometric analysis described in this section has been made publicly available through *Zenodo* (DOI: 10.5281/zenodo.4768544). The liberal license (CC BY 4.0) allows the code to be adapted at will or used for one's own research purposes, provided that an attribution is given. The code has the following major dependencies: Python 3.6+, NumPy, scikit-learn, and SciPy.

[18] We used the Ward clustering algorithm, an agglomerative clustering method that allows for each pair of clusters with a minimum distance to be merged.

[19] Similarly, we developed dendrograms for various amounts of extracted features, ranging from 500 to 2,500 most frequent words. The different plots resulting from this experiment stayed more or less the same, suggesting that the obtained result is a robust one.

[20] More specifically, we ignored words that appear in less than five novels (i.e., character names and highly novel-specific terms).

[21] "De vraag is of je het kind helpt door het zo'n verhaal aan te bieden. We geloven direct dat Guus Kuijer daarin integer is, maar de vraag is met die opmerking natuurlijk nog niet beantwoord. Wat moet het lezertje met deze droom aan in het dagelijks leven?"

Works Cited

Armstrong, Thomas. *The Human Odyssey: Navigating the Twelve Stages of Life*. Ixia, 2019.

Berglund, Karl. "The Space Between: The Usefulness of Semi-Distant Readings and Combined Research Methods in Literary Analysis." *Digital Humanities in the Nordic Countries: 2nd Conference*, Uppsala U, Sociology of Literature, 2017.

Brennan, Timothy. "The Digital-Humanities Bust." *The Chronicle of Higher Education* 15/10/2017. www.chronicle.com/article/The-Digital-Humanities-Bust/241424.

Cross, Amy, Cherie Allan, and Kerry Kilner. "Digital Curation, AustLit, and Australian Children's Literature." *International Research in Childrens Literature*, vol. 12, no. 1, 2019, pp. 1–17.

Da, Nan Z. "The Computational Case against Computational Literary Studies." *Critical Inquiry* vol. 45, no. 3, 2019, pp. 601–39.

Daelemans, Walter. "Explanation in Computational Stylometry." *Computational Linguistics and Intelligent Text Processing*, edited by Alexander Gelbukh, pp. 451–62, Springer, 2013.

de Vries, Anne. "Wat zegt de kinderjury? De stem van de kinderen bij de canonisering van kinderboeken." *Het paard van Troje. Niet-schoolse teksten in het onderwijs*, edited by Helma van Lierop-Debrauwer et al., NBLC, 1996, pp. 107–11.

Fish, Stanley. "The Digital Humanities and the Transcending of Mortality." *The New York Times* 09/01/2012, https://opinionator.blogs.nytimes.com/2012/01/09/the-digital-humanities-and-the-transcending-of-mortality.

Fitzsimmons, Rebekah, and Suzan Alteri. "Possibly Impossible; Or, Teaching Undergraduates to Confront Digital and Archival Research Methodologies, Social Media Networking, and Potential Failure." *Journal of Interactive Technology and Pedagogy*, vol. 14, 2019, https://jitp.commons.gc.cuny.edu/possibly-impossible-or-teaching-undergraduates-to-confront-digital-and-archival-research-methodologies-social-media-networking-and-potential-failure/.

Giddens, Eugene. "Distant Reading and Children's Literature." *The Edinburgh Companion to Children's Literature*, edited by Clémentine Beauvais and Maria Nikolajeva, Edinburgh UP, 2017, pp. 305–7.

Green, Lorraine. *Understanding the Life Course: Sociological and Psychological Perspectives*. Polity, 2010.

Hutcheon, Linda. *A Theory of Parody: The Teachings of Twentieth-Century Art Forms*. Methuen, 1985.

Hockey, James, and Alison James. *Growing Up and Growing Old: Ageing and Dependency in the Life Course*. Sage, 1993.

Holmes, D. I. "The Evolution of Stylometry in Humanities Scholarship." *Literary and Linguistic Computing*, vol. 13, no. 3, 1998, pp. 111–17.

Jänicke, Stefan, et al. "On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges." *Eurographics Conference on Visualization (EuroVis)—STARs*, edited by R. Borgo et al., The Eurographics Association, 2015, doi:10.2312/eurovisstar.20151113.

Jockers, Matthew. *Macroanalysis: Digital Methods and Literary History*. U of Illinois P, 2013.

Jockers, Matthew L., and David Mimno. "Significant Themes in 19th-Century Literature." *Poetics*, vol. 41, no. 6, 2013, pp. 750–69.

Joosen, Vanessa. *Adulthood in Children's Literature*. Bloomsbury, 2018.

———. *Connecting Childhood and Old Age in Popular Media*. U of Mississippi P, 2018.

Juola, Patrick. "Rowling and 'Galbraith': An Authorial Analysis." 2013, https://languagelog.ldc.upenn.edu/nll/?p=5315.

Karsdorp, Folgert. *Retelling Stories: A Computational-Evolutionary Perspective*. Radboud U, 2016.

Kessler, Jason S. "Scattertext: A Browser-Based Tool for Visualizing How Corpora Differ." *ArXiv:1703.00565 [Cs]*, 2017, http://arxiv.org/abs/1703.00565.

Kestemont, Mike. "Stylometric Authorship Attribution for the Middle Dutch Mystical Tradition from Groenendaal." *Dutch Crossing: Journal of Low Countries Studies*, vol. 42, no. 3, 2018, pp. 203–37.

Kestemont, Mike, and Dirk Van Hulle. "Periodizing Samuel Beckett's Works: A Stylochronometric Approach." *Style*, vol. 50, no. 2, 2016, pp. 107–202.

Kirsch, Adam. "Technology Is Taking Over English Departments. The False Promise of the Digital Humanities." *The New Republic*, 5 May 2014, https://newrepublic.com/article/117428/limits-digital-humanities-adam-kirsch.

Kuijer, Guus. *De jonge prinsen*. Querido, 1986.

———. *De man met de hamer*. Arbeiderspers, 1975.

———. *De redder van Afrika*. Arbeiderspers, 1989.

———. *De tranen knallen uit mijn kop*. Querido, 1980.

———. *De zwarte stenen*. Querido, 1984.

———. *Drie verschrikkelijke dagen*. Querido, 1976.

———. *Eend voor eend*. Querido, 1983.

———. *Een gat in de grens*. Querido, 1975.

———. *Een hoofd vol macaroni*. Querido, 1979.

———. *Florian Knol*. Querido, 2006.

———. *Grote mensen, daar kan je beter soep van koken*. Amsterdam: Querido, 1976.

———. *Het boek van alle dingen*. Querido, 2004.

———. *Het dochtertje van de wasvrouw*. Querido, 1980.

———. *Het geluk komt als de donder*. Querido, 2000.

———. *Het is fijn om er te zijn*. Querido, 2000.

———. *Het land van de neushoornvogel*. Querido, 1985.

———. *Het vogeltje van Amsterdam*. Arbeiderspers, 1992.

———. *Hoe Mieke Mom haar maffe moeder vindt*. Querido, 1978.

———. *Ik ben Polleke, hoor!* Querido, 2001.

———. *Izebel van Tyrus*. Arbeiderspers, 1988.

———. *Krassen in het tafelblad*. Querido, 1978.

———. *Met de poppen gooien*. Querido, 1975.

———. *Met de wind mee naar zee*. Querido, 2001.

———. *Olle*. Querido, 1990.

———. *Op je kop in de prullenbak*. Querido, 1977.

———. *Papa is een hond*. Querido, 1977.

———. *Rose, met vrome wimpers*. Meulenhoff, 1971.

———. *Tin Toeval en de kunst van het verdwalen*. Querido, 1987.

———. *Tin Toeval en de kunst van Madelief*. Querido, 1989.

———. *Tin Toeval en het geheim van Tweebeens-eiland*. Querido, 1987.

———. *Tin Toeval in de onderwereld*. Querido, 1997.

———. *Voor altijd samen, amen*. Querido, 1999.

Moretti, Franco. "Conjectures on World Literature." *New Left Review II*, no. 1, 2000, https://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature.

———. *Distant Reading*. Verso, 2013.

Neiman, Susan. *Why Grow Up? Subversive Thoughts for an Infantile Age*. Farrar, Straus, and Giroux, 2014.

Owens, Trevor. "Defining Data for Humanists: Text, Artifact, Information or Evidence?," *Journal of Digital Humanities* vol. 1, no. 1, 2011, http://journalofdigitalhumanities.org/1-1/defining-data-for- humanists-by-trevor-owens/.

Rosen, Jeremy. "Combining Close and Distant." *Post45*, 2011, http://post45.research.yale.edu/2011/12/combining-close-and-distant-or-the-utility-of-genre-analysis-a-response-to-matthew-wilkenss-contemporary-fiction-by-the-numbers/.

Stephens, John. *Language and Ideology in Children's Fiction*. Longman, 1992.

TEI Consortium. *Guidelines for Electronic Text Encoding and Interchange [Version 3.5.0]*. 2019. http://www.tei-c.org/Guidelines/P5/.

Tromp, Herman. *Over de jeugdboeken van Guus Kuijer*. Arbeiderspers, 1982.

Ueno, Miki. "Computational Interpretation of Comic Scenes." *Distributed Computing and Artificial Intelligence, 13th International Conference*, edited by Sigeru Omatu et al., pp. 387–93, Springer International Publishing, 2016.

van de Ven, Inge, et al. *Analyzing Reading Strategies: Bridging the Gap between Close and Distant Reading*. 2017. Paper presented at DH Benelux, Utrecht, Netherlands.

Ward, Joe H. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association*, vol. 58, no. 301, 1963, pp. 236–44.

Zhang, Yin, et al. "Understanding Bag-of-Words Model: A Statistical Framework." *International Journal of Machine Learning and Cybernetics*, vol. 1, no. 1, 2010, pp. 43–52.