

Sara Schwebel: Thank you for joining us for the Center for Children’s Books 2020-21 Speaker Series! We are delighted that the move to a virtual platform enables people to gather from afar.

A few logistics:

- Please complete the survey to help us understand who is gathered in the audience today.
- If you run into technical problems, please reach out to the iSchool Help Desk (see chat for details)
- When we arrive at the Q&A, please **type** your question or comment for the speaker in the chat box, and we’ll work our way through them

Without further ado, I’m pleased to introduce today’s panel of speakers, collectively known as the Data-Sitters: Lee Skallerup Bessette (Georgetown University), Katia Bowers (University of British Columbia), Maria Cecire (Bard College), Quinn Dombrowski (Stanford University), Anouk Lang (University of Edinburgh), and Roopika Risam (Salem State University). Today’s panelists are experts in a range of disciplines—from English Literature to Slavic Studies, literature in translation, secondary education, and academic technology—but share in common an expertise in Digital Humanities and a passion for “The Baby-Sitters Club” series fiction. *The Baby-Sitters Club* was a wildly popular Scholastic series published between 1986-2000 that sold a stunning 176 million copies (by comparison, the Harry Potter books had sold just slightly more in the US, 180 million copies, by 2018).

Following in the footsteps of an earlier generation of feminist scholars who brought attention to the ghost-written Nancy Drew books of the first half of the twentieth century, the Data-Sitters are exploring a girls’ series that flourished during the age of suburban malls and big box stores. The series may not be known for its high literary value, but it provides intriguing windows into gender, social class, and race at a time when “multicultural” representation was valued but the OwnVoices movement lay in the distant future.

Yoking their interest in the texts of their adolescence with current DH methodologies, the Data-Sitters are asking what can be gleaned from study of the Baby-Sitter books as a corpus—in English and in translation—and what they can teach others about DH methodologies in the process by publicly documenting their work, inclusive of all fits and starts. As the Data-Sitters website makes clear, they’re having a lot of fun doing both!

I’ll now turn over the mic to them. Their talk is titled, “Tropes and Tribulations: Exploring Computational Text Analysis with the Data-Sitters. Welcome!

Quin Dombrowski: There we go. Sorry, sharing the screen always screws everything up. Hi everyone, and thank you so much for joining us today. We are the Data-Sitters club. And as you can see, our wonderful book cover here was created by Claire Chenette.

And yeah, we’re here to talk about this project that we’ve been working on now for a little bit over a year. And yes. So, like many stories, it started in Las Vegas. I was going on vacation for the first time in about four years, having some drinks, and I saw an article it’s been written,

and posted online that included an interview with Shannon Supple, who is the curator of the Ann M. Martin papers at Smith College.

I promised my husband that I would not do any work on this vacation. And as my job is non-English digital humanities, I thought to myself “How better to pass the time in Vegas, you know, while drinking and sitting there with my laptop, then to try to do some digital humanities work in English?” And this article about the Baby-Sitters club and Ann M. Martin, you know inspired me to start looking for, you know, what if I were to put together a corpus of these books that I had these fond memories of reading as a child. And so, I did so.

And you know, like much of digital humanities work, the kind of the social life of DH takes place on Twitter. And so I started tweeting about it, and I quickly discovered that a number of people I knew were, in fact, you know, fellow readers of the Baby-Sitters club and in fact Baby-Sitters Club super fans. And before long, we decided we would get together and actually do something with this. So, I mean, part of the goal of the Data-Sitters Club is, you know, really an act of feminist pedagogy. There's a lot of, you know, very impressive work in computational digital humanities, but it usually draws upon corpora that are less familiar to people, less relatable. And, you know, kind of, written by and for kind of the usual suspects of computational, text analysis which tends to be white guys.

So, being able to you know put together a project that uses material that for, you know, a group of people who are not necessarily as well served by the materials that are available. You know, using analogies that relate to things like child care. And, you know, shopping sprees and things like that to explain, you know, concepts of machine learning. So, writing these things up and making kind of the work that we do in computational text analysis and computational DH more accessible is one of the driving forces of this project and also of course yes, having fun, as we do so.

Right, take it away. Katia.

Katia Bowers: Hey. Oh, so I am Katia, I am an associate professor of Slavic studies, and I'm going to talk a little bit about our corpus. But before I do that, it might be helpful to understand what these things are that are on the right side of the slide. [Slide titled: The (anglophone) corpus] The Baby-Sitters Club is, as Sara said, a series of books that were widely popular. And what we do in the Data-Sitters Club is... So, the overall idea is to create a colloquial, comprehensive digital humanities guide to text analysis, and each of the things that you might be wanting to learn through this is described in what we call a book.

So, Quinn and I wrote together book number two, Katia and the Phantom Corpus. And this describes how we created our corpus which I'm going to talk a little bit about now. So, The Baby-Sitters Club, Sara gave a good introduction to this already. It consists of a vast number of books and various sub series. So, there's the main series, which is 131 books. And then there's a number of other series, you can see them here. So, there's super specials, mysteries, super mysteries, Special Edition reader requests. And there's only like three of these, two of which are about one of the character's boyfriend. Portrait collection. And then there's two spin -ff series: Friends Forever and the California Diaries. And now they're also publishing graphic novels, which are excellent. And there is also a prequel novel. So, you can get a sense of that

So, our corpus consists of two main parts. We have the Anglophone corpus and we also have the translation corpus. The translation corpus is made up of over 200 books and you can see some of what this looks like here. [Slide titled: translation corpus] So we have nine different languages. And in addition to that, sort of books that are subdivided into categories of books, that

are subdivided into different languages. So our Francophone corpus are 144 books, and that divides into a group of books written in Belgian French from the '90s, French French from now and French French from the '90s, Quebecois French from the 90s, and then in addition we have graphic novels in French French and Quebecois French. And so, these are used in our own sub series in the multilingual mysteries, which Quinn and Lee have been working on. Next slide. And over to Maria.

Maria Cecire: Hi everyone, I am I suppose the resident Children's Lit scholar, which I really enjoy. I love bringing my history of children's literature and theories of childhood to what we do in *The Data-Sitters Club*. And my, I have a book about children's literature, called *The Enchanted: The Rise of Children's Fantasy Literature in the 20th Century*, which came out in 2019. And so, as you can see it's not on *The Baby-Sitters Club* or the '90s girl realist series or anything like that. But I read the books as a kid also and Quinn invited me to join the team around the time the book came out for my insights into the field. And because I have some background and exposure to digital humanities methods through my work with the Center for Experimental Humanities which I directed at Bard College, until just recently.

So, you know, it's been a lot of fun, but sometimes my role as the person coming most centrally from the field of children's literature studies can make me a little bit of a killjoy. And this really comes through in my book, *Data-Sitters Club* number three, *The Truth About Digital Humanities Collaboration*. And I'll give you a little overview of what happened there and it'll also give you a sense of the way I'm thinking often about the questions that we address together, and then how we can sort of start working together as a group to sort through the different approaches to these questions together.

So, for example, Quinn and the other members of the team were doing text comparisons with the newly OCR corpus that Katia just told us about, and all these enthusiastic emails were coming into my inbox about the content changes between the original and rereleased books. So, some of these changes were things like swapping 21st century technologies in for ones from the 80s and 90s or fashions. So thinking about taking out references to permed hair or something like that or taking out VCRs, or changing money or to adjust it for inflation those kinds of things. So everyone was like getting really excited about it and there was a whole flood of emails about this, thinking what can we do with it, how should we present this information? But I knew, and many of you will know, that updates to rereleased additions for children's literature is pretty standard fare. It's pretty common for a couple of reasons, and one of them is the moral responsibilities that are associated with giving children books. Which, typically since the late 18th century, requires ensuring that their contents align with contemporary values that are considered appropriate for children. And also children's literature's status as a commodity, and one for that was for long-time dependent on slow burn sales over generations, so often required updating for rereleases, to make them relevant for new generations of sales.

So here are a few examples of some of the more morally motivated changes that have been made to children's literature [slide titled: What are the limits of computational methods?], and maybe one of the most iconic examples is the very designation of the Grimm Fairy Tales as Children's Household tales, as they were called children's household tales. But they were originally collected as the logical examples, and for the preservation of folklore, but they were quickly picked up as children's literature and then that immediately led to outcries about their inappropriate material. And that led to very swift broader realization and adjustments to the books to make them more palatable for children in the 19th century, which of course even now

we talked about Grimm's as being really grim. But these versions are often the versions that were themselves kind of cleaned up in that period.

More recently, we have seen in the '60s and '70s, the transformation of the oompa loompas in *Charlie and the Chocolate Factory* from quote unquote "amiable black pygmies" from Africa to whitened figures from the imaginary Loompa Land, and that was in response to changing norms around racial representation in the '60s. And you can see in the '70s and that image there is taken actually from Phil Nel's thoughtful reflections on his blog "Nine Kinds of Pi", which many of you may know, but if not the URL is in the corner of the screen (<https://philnel.com/2010/09/19/censoring-ideology/>).

And very, very recently, you probably heard about the outcry following the Dr. Seuss enterprises' decision to stop publishing six of Seuss' books that contain racist representations and Phil again, who has a book on this, *Was the Cat in the Hat Black?* In this article that you can see on screen here discusses this decision as a moral decision. And it's interesting here, because the way he talks about it kind of spans the two things I'm talking about here moral implications and commodities. He argues that Dr. Seuss enterprises "has made a moral decision of choosing not to profit from work with racist caricature in it and they have taken responsibility for the art they're putting into the world" and I would support that analysis. And he refers to the books, as a recall, and in that sense, referring to, you know, these books as commodities that need to be recalled. And that helps us then to think about the other reasons that children's books are constantly being updated and changed.

And so, you know, we see this happen all the time. So for example, you know, Enid Blyton's Famous Five books were we released in around 2010, and *The Baby-Sitters Club* books were also rereleased around that same time. And part of this is because of what I mentioned before. Up until really Harry Potter, children's literature sales were predominantly reliant on this idea that they would sell over multiple generations and that people would buy books that they had enjoyed themselves as children for the next generation. And so, rereleases are not unusual. We saw this recently with series like *Goosebumps* as well.

And if you jump to the next slide please. We will see even that, in things like the New York Times announcement about *The Baby-Sitters Club* being rereleased they explicitly describe that some of the references acknowledging fashions are going to be updated. So, all this brings us back to this moment of email consternation for me, where I sat there with all this information and history that I knew meant that computationally discovering these differences in the rerelease books was actually not that big of a deal. So, it created this kind of uncomfortable collaboration moment, you know. What do I say? how do I stop everyone's enthusiasm, but also you know I think it's important that they know how the field would see this and how we might treat it from the perspective of children's literature and childhood studies.

So, I tried to really gently break it to them and at first it was a little bit awkward, but everyone was really, you know, very cool about it. And that quickly led to a meta conversation about collaboration and, you know, how do digital humanities, techniques, how do they work in, you know, in ways that might be strengthened by being really closely aligned with people's work in the field. And also vice versa how might work in particular fields be strengthened by having people from digital humanities help us to think about our questions from new perspectives. So that's, you know, something that I ended up the then writing about really that collaboration and experience in my book. And by my book I mean my *Data-Sitters Club* book, which is a blog post. And from then we've been having lots of ideas brewing and bouncing around.

I'm not gonna read all this but this is a quote from Peter and Wendy, the novelization of J. M. Barrie's Peter Pan, which I was using to kind of try to explain my interest in how we might digitally map conceptual worlds and particularly the conceptual worlds of children's literature and specifically The Baby-Sitters Club. By looking at our corpus and I was like, how might we do this you know I was interested in this idea of like the constraints of a world, and the shape of a world, how children are represented as thinking about their world in children's literature, or in this case, adolescent girls. And so, Anouk and I kind of bounced some ideas about it, and she of course, took all of her wisdom and expertise and went off running in a really exciting direction, so I will pass it over to Anouk.

Anouk Lang: Okay, thank you. I'm Anouk, and I'm a senior lecturer in digital humanities, English department at the University of Edinburgh. I'm going to turn my camera off, please excuse me for doing that. So yes, Maria set this up really nicely, and this really gorgeous quotes from Peter and Wendy I will just keep on the screen, as I talk to you a little bit about the kinds of computational work that I did to investigate this really interesting idea of the map, of how you might map the conceptual world of the series that stretches over, you know, as you've heard from Katia, so many different books and also kind of sub series within the main series.

So, the Data-Sitters Club books or other posts that I've written or co-written so far are about computational text analysis. And rather than kind of retread that territory, if you want to read them you can go into the site and read the books. Instead, I want to talk about a form of machine learning analysis that has been causing some excitement in digital literary circles. In the past 25 years, which is called Word Vectors, also known as word embedding, And to when you work with these you're essentially creating a mathematical model of the relationships between words which you can query. And it can give you a sense of the more subtle associations, such as like analogies or oppositions that are working in the discourse across a corpus.

And the background to this is that in my actual area of specialization, which is the 20th century, I've used word embedding to explore how big corpora conceptualized place in ways that are different from the habitual taxonomies that we use. So we've got things we've got concepts like north, south, east and west and we divide places into things like cities and states and villages, but literary takes don't talk about place in those ways. Instead, it's much closer to what you've got an example of on the screen here. So kind of a map of concepts and ideas and images that aren't linked in any kind of concrete way but instead in kind of mobile and dynamic ways. And then you've got kind of other ideas, other maps sharing consistently through.

So, just inspired by, you know this quote and also kind of discussions we've been having, and that's kind of came off our last paper to the Stanford Lit Lab about mapping the conceptual worlds of The Baby-Sitters Club. But I wanted to use Word Vectors to probe the limits of the books' spatial imaginaries, with a look at how The Baby-Sitters Club might differ from comparative texts, which I think is these shown by these quotes. Now, if you've read the books, you'll know that they're set in the fictional town of Stony Brook, which is in Connecticut. So, what I did is I train the model on our 179 Baby-Sitters Club texts and that's roughly 4.7 million words. And I asked it for the 10 closest sectors to Connecticut, so that was that first box that came up. And you can see what happens when you do that. You get a list of the names, so they are mostly towns and cities, you got a few states in there. The word Jersey, New Jersey, New York to New York. England is in there.

So you can think of vectors as the cloud of semantic associations that surround a word. So, these lists are not delivering anything unexpected. You know that if Connecticut is a place

name and its associated with other place names, so that might not seem very interesting. But it does get more interesting, because as you iterate through that initial list of vectors that you got and you look for the vectors associated with those. What you see is the same place names coming up again and again, so sure there are some new ones there's kind of gradually get places bit further afield like England and France, and you get a very few words and other classes like homesick or ER visits. But whereas in other corpora where I've done this kind of testing investigation, they tend to take you out to other place names. They tend to take you out to other conceptual terrain, but here they just keep circling back to the same fairly contained list of cities and states on the two coasts.

And so, what I've done on this next slide is just highlight them so you can see like how often they recur. So, you start with Connecticut, and you get this list, California, Berkeley and then you can see that if you if you search for California you get some of the same words coming up. So you can see how they recur. Of course what you need to do is to check the context in which these appear, so the word England might, of course, sometimes be being used in the phrase New England, and this is a screenshot from a concordance and conch just one of the tools I wrote one of the BBC books on. And so that is quite useful for bringing up all the instances of a particular term or phrase and so if you, if you look England up using that you can really quickly see that New England accounts for only about a quarter of those instances.

So just funny and quickly, it's one thing to investigate how the places are configured in our corpus, but it's also important to understand how these associations are working in comparison with other texts or the corpora, And this is where a project called YAMS comes in and Quinn's gonna say more about this later, so I won't steal her thunder. But this is just kind of give you a foretaste of how a logical opposite of young adult fiction might be valuable. So I ran the same queries on a sub corpus of roughly comparable YA books so also about 7 million words, mostly series, mostly published in the 80s and 90s. And you can see the difference in the kind of results that it returns. When you start looking for, if you like the edges of it spatial imaginary. So if query this corpus for California you end up with a much broader geographical reach, like you're immediately out in places like Paris and Africa, and you've also got other terms coming into the word community or descriptive terms. You've got inland and wooded when search for Manhattan, I think.

So, this isn't a like for like comparison because I had to use nine different series in order to make something as big as the BSC corpus. But the variety here does align fairly well with what I found when I've looked at place in more homogenous corpora. And it can give you a sense of the way that the geographical imaginary of the Baby-Sitter's Club books is configured, and also the way that it's limited, so I'll pass over to you I think it's really Roopika next.

Roopika Risam: It is me next Hi, I'm Roopsi. I am the resident computational textual analysis skeptic. My complaints are well documented in my book, *New Digital Worlds*, but you know my background is in Postcolonial and African Diaspora studies, so I'm really concerned with what doesn't make it into a corpus and how do the broader omissions of literature by black, brown indigenous people in publishing and digitization all then end up impacting the kinds of claims that we can make about literary history with computational textual analysis. And the ways they can kind of overwrite the contributions of black, brown and indigenous writers.

So, in light of this, it may seem weird that I would get involved in a project like this, but I do have actually an encyclopedic knowledge of what happened in every single Baby-Sitters Club book. And that's a little bit creepy sometimes. And so, it's really been an interesting opportunity.

And, you know, actually just sort of falling a little bit on what Maria had been saying, it's really shown me what collaboration can do, particularly and how collaboration can make us think differently about our practices in digital humanities. So I just want to highlight one moment from our work, related to the, the book number five, *The DSC and the Impossible TEI Quandaries*. And that was a very collaborative book. So collaborative I guess none of our names are on the cover, but we were joined by Elisa Beshero-Bondar, TEI expert.

And so, in case you're not familiar with TEI, it stands for text encoding initiative, and it's a semantic markup language standard that helps us digitize texts and look. You can markup specific features of a text with tags that look like HTML tags, but they don't actually render anything on a web page but they indicate locations in particular features in a text. So in a poem for example you could mark where stanzas are. and I, I am as skeptical of TEI, as I am of computational textual analysis, more generally, but I do have a TEI digital edition of poems *Harlem Shadows* which I created with Chris Forster when we were both in graduate school, and it's actually, you know people use it in their classes, people have written about it, it's spawned little sister projects, which is very appropriate for Baby-Sitters Club

But other than that, for me that exercise and doing it felt like kind of pointlessness and an exercise and doing a TEI edition just to say I had done a critical digital edition and get my street cred. And you know if I'm going to be a digital humanist and I don't do TEI and I have a PhD in English, what does this all mean? So when we were working on this TEI book, I decided to take the opportunity to interview, Elisa Beshero-Bondar, TEI expert, and really try and understand what the big fuss about TEI is all about and Katia came to the interview as well, to co-interview Elisa with me. And it's actually pretty funny because I repeatedly asked for, Elisa if TEA is a cult. But what I actually, I started to understand what the fuss about TEI actually was. And she convinced me to think about how TEI can turn a text into a data treasure chest, which is literally the words she used that can then be queried.

So you could mark up all the parents is of a character all the particular locations of the text, and then with regular expressions query that data and extract all the appearances or count things or do whatever else you want to do that's fun with computation technical analysis. And, while I still find TEI to be wildly time consuming, I really did come away with this completely different appreciation for what it could do beyond, "all right I made a digital edition maybe somebody else to go do something with it that day." And, you know, that sort of also that moment of, you know, being able to be in collaboration with somebody who really really understood the possibilities and was also able to be alongside Katia and I thinking through our questions which in some cases was trying to push against what TEI even is. So I was like, we started talking about querying the TEI right, and so that was just really, really eye opening about what it could do.

And this collaboration, I think, with the Data-Sitters Club has really for me marked the beginning of far more collaborative possibilities in my own research agenda. Well, not just my research. Quinn and I were both it turns out accidentally running for vice president of the Association of Computers in the Humanities and, you know, it was a little bit of a kind of an awkward situation, but we decided to make it not awkward and then eventually we decided that we should just both be vice presidents and who cares who voted for anyone? And that is what we did. And so I really feel like, you know this is a such a great model of collaboration that has expanded beyond what we do and really shown the value and utility of computational textual analysis as a method specifically when you're working with a team that's comprised of enough of the right kind of expertise. The children's lit expert, the word vector expert. The Quinn. The

Katia. The Lee. So all these people bringing their expertise together to really realize a kind of exciting outcome.

Katia Bowers: Hey I'm back. So, as Roopsi was saying, we all have our specific kind of roles in the club, although none of this is of course formalized, and my main role is as kind of the newbie. So I should say that I met Quinn at a Slavic DH workshop. And I was inculcated to use Roopsi's cult analogy into the Data-Sitters Club when I was trapped in the Newark Airport for many hours afterwards, and was on my phone on Twitter. And at the time when I was brought in, I really had not done a lot with digital humanities at all. I can do tech stuff but I had no idea how to put all this together. And so working collaboratively with the Data-Sitters Club has really been a great opportunity or experience for me to both feel like I have a place as a fairly new person because I'm usually the person who asks a lot of questions like why are we doing this or like, how does this work and points out when kind of we've skipped a step doing something that we should be explaining. And this also has allowed me to get excited about new things so I actually, I'm not sure I can call myself a new digital humanities person anymore because I actually proposed a book to Quinn and signed up for a workshop to figure out how to do the thing and it's all happening.

So, I thought it would be helpful to discuss a little bit about how we come up with these books, and what they do, and the way I'm going to do that is by talking about the book that I wrote by myself: *Voyant's Big Day*. So, the book is about *Voyant*, and *Voyant*, for those of you who don't know, is an online suite of tools that you can load a text or corpus into, and then you can use it to do different visualizations of the text or corpus. And you can do analysis based on those visualizations. So, the task that I had set out was to use *Voyant* tools to write a book about using *Voyant* on tools for newcomers to *Voyant* tools, and I myself was a newcomer to *Voyant* tools. Although I would hazard a guess that many newcomers to *Voyant* are people who are just taking first steps into DH like I was and are also encountering *Voyant* tools as kind of the first thing. So the goal for this was to really demonstrate all the things that *Voyant* can do.

So, the key need in writing any the *Data-Sitters Club* book and this usually comes from a discussion with the group. This kind of collaborative discussion is to figure out a question to ask or a question to kind of frame the post. And of course the best question for this book was drawn from my childhood reading habits which, like Roopsi's, heavily focused on the *Baby-Sitters Club*, namely, what's up with the weird baby sitters club slang? So, when we first started the club, we were actually making fun of this slang at meetings. It's a really memorable thing about the series, but when I was rereading the series as an adult, which started also after I joined the *Data-Sitters Club*, I began to notice that the slang just wasn't there. And in fact, I found the first instance of saying only in the 36th book *Jesse's Babysitter*. And when that slang, which is fresh, which means cool, was used, the narrator explained that it was a word that the *Baby-Sitter's Club* used all the time, but obviously it hadn't been used in the previous 35 books.

And so, we decided to use *Voyant* to figure out what was going on with the BSC slang and this is a pretty good use of *Voyant* because we *Voyant* is about counting words and seeing how they fit into things and seeing how they fit across the corpus. And this is exactly what I needed to do with the slang. Now I should say that this was useful to look at *Voyant* in a deliberate way right to kind of like, create a book out of, but I surprised myself by finding something potentially meaningful. So, one of the most ridiculous *Babysitters Club* slang words is *dibble*, which is short for *incredible*, and it means cool. And this red spike in the middle is what I found. So this red spike in the *Voyant* trends window denotes all of the *dib** appearances

in Baby-Sitters Club number 39, Poor Mallory, and it sticks out against the field of other slang words that I was looking for. It's also, I should say Dibble is a particularly useful babysitter slang word for these purposes because it's the only babysitter slang word that isn't also a different word. So a lot of the other words are like fresh or stale or distant or chilly and they can be used in other contexts but dibble is really only used in a slang context.

So when I cut out to read the rest of the non-dib* words, I could see the dibble really only appears in five books which were all published between November 1990 and May 1992. Doing some additional investigation allowed me to make a conclusion, namely really only one character is using dibble in the books, and that's Mallory. And Mallory claims that the other babysitters are saying it all the time but in practice, they are not actually doing this. And the TL;DR version of this is, although the Baby-Sitters Club fans claim Baby-Sitters Club slang is memorable. It only really appears in the books rarely and it was really mainly only during the early 90s, and although the Baby-Sitters Club describes a vibrant lexicon of words that mean cool, in the end, they mostly just say cool to me.

Quinn Dombrowski: All right. Um, so speaking of memorable features of the Baby-Sitters Club books, another of these is the infamous chapter two. And for those of you who've read these books, it's made fun of constantly. The fact that chapter two reliably gives the whole backstory of the club, plus the history of all the characters, like in every single book. But like, it's always there. And so, we thought that would investigate this in a couple different ways, and one of the ones that we use was looking at sequences of 6-grams. So, some of these are like very clearly these tropes like “since we use her room” and this is a justification for why Claudia is vice president, “long black hair and almond-shaped eyes” also serves as a racialized trope for Claudia, “who was Vietnamese was adopted” we're talking about Christie's adopted younger sister, “move to California got married”, and we're talking about Dawn's family, “became best friends, now they're stepsisters.”

So those show up a lot repeated across all these chapter twos. We also get some noise like “I didn't tell you that” and “trying to line up a sitter for” that's like, congratulations that's colloquial English. But still, there's no reason to think those like colloquial Englishisms are distributed evenly so we thought we would go with this as a metric. And we've looked into chapter twos with high overlap. So these are all chapter two that had a very large number of these 6-grams. I'm overlapping between them you having the same phrases repeated. And at first, we thought we would color code by narrator. You know, maybe the different narrators emphasize different things you know talking about Mallory versus Claudia's family, things like that. But what you get is like, you know, a fruit salad that's like very well mixed, so like it's not really cohering this way. Unsurprisingly, but like sort of fun to prove, if you re-color it by ghost writer, you see these much clearer clusters here, where like once you've written one chapter two, your other chapter twos odds are, they are going to be a lot like it because it's the same stuff repeated over and over again.

We also tried a couple other different metrics we got into, exploring some of the challenges of using algorithms and you don't really understand what all the parameters do. Hilarity ensues and so forth. But this is another thing that we explored for the project.

Lee Skallerup: Awesome. So, I'm kind of like Dawn in the books, but instead of coming later in the series or a few books into the series from California, I'm coming from DC via Montreal, Quebec, Canada. And so, what was really exciting is they found all of these different translations

into the same language into French, and said hey I wonder who can tell us about those things and I just so happen to have a book on exactly that except not from English to French but from French to English. My area of expertise is actually French to English, English to French translation particularly but not exclusively within the Canadian and Quebecois context. So Roopsi, having heard me drunkenly talk about the history of Quebec and translation and language politics on more than one occasion, was like, oh I seem to remember Lee knows something about this, so I was brought in and asked to bore everyone with long histories of Quebecois politics and linguistics and publishing history, which turned into multi lingual mysteries.

So we first started by saying, Okay, how do we find the metadata for all of these? Who translated what and when? When were they published? Where were they published? Within those kinds of things and so this was me scraping various national library sites, and then using DH techniques to clean up that data to be able to create a readable spreadsheet of what's been translated, what's there, what are they called, who translated it, when were they translated. And so we do have three different. ones actually four different ones if we want to count that, so France was translated in the 90s, but they also have a what they would call vulgarized version that came out more recently. So the original translations were sort of translated in high French you could say, and then the more recent versions of the translations are translated in a more simplified, vulgarized French. France's translations involved American girls babysitting American kids in America so they just translated it. They didn't bother adapting it in any way, shape, or form, which is not unusual for translations.

But certainly in children's literature, there is the possibility of something like Quebec, which adapted and adopted the books where it is Quebecois girls except for Stacy who's from Toronto, which makes sense as that's the kind of New York City of Canada, babysitting Quebecois kids. Except for the movie star who does actually go to California, which is a whole other thing. And then in Belgium, translated about the same time as the Quebecois one. You had French girls, so French from France, not Belgian girls, French from France girls, except for Jesse who is from Burkina Faso babysitting French kids in France. And so what's fascinating about these is that, much like Maria saying it wasn't news that there are adaptations to be made, it's not so much that there were adaptations, that's the interesting thing. It's what were the choices that the translator made for their particular target audience, for that particular time in how they chose to translate this American, this quintessential American suburban girl experience. So how is that done for the various foreign audiences?

And it's interesting that you know the distances and all of that and one of the more interesting examples that we're currently exploring is food. And so one of the, one of the most. I wouldn't say egregious but certainly interesting translation choices where we have spaghetti O's, which is a you know quintessential American, kind of, pasta in a can. But it's translated in it the Belgian one as Osso Buco, which is an intricate, elaborate sort of meal that in no way is resembles opening a can and plopping it in the microwave. So there are all of these cultural things between books and games and food that is just fascinating to see and what does that say about what assumptions are being made about the target audience that they're translating for. So, just a lot of great things about what the translation say about the target culture and the target culture's interpretation of the specificity of the Baby-Sitters Club.

Here's some of the wonderful adaptations of the cover art over time. The one in the lower right here is representative of some of the recent re-translations and additional translations from the Netflix series.

Quinn Dombrowski: Alright so thinking more broadly about children's literature, you know, like the Baby-Sitters Club is great but I mean what can projects like this teach us about children's literature? And I think one of the things that is worth thinking about is really looking beyond prestige examples. When we talk about kind of trends in children's literature, like we think of the classics often like the award winners. But there is so much more out there that, kids and young people are actually reading. There's the pass around book that like, you know, no one wants their mom to catch them having and throw away books like the Baby-Sitters Club, where like, you know, there are people who kept like sort of precious collections. There's also many series that are like the Baby-Sitters Club that people were not as emotionally attached to and like you can now find for like 50 cents on AbeBooks or whatnot.

So, you know, one of the ways that I've been working on expanding out from the Baby-Sitters Club is this project that we're calling YAMs, along with Nicole Nomura and Jennifer Wolf at Stanford. Currently we have about 15,000 the middle grade and YA novels, mostly from the 2000s to present, but some you know dating back to the, the 80s and 90s and I started collecting some wonderful things like Samantha Slade, Confessions of a Teenage Frog, which is basically like an Addams Family meets Baby-Sitters Club but like rip off, it's fantastic. You know, and so we have all these books and we have some questions that we're trying to ask them, you know, like how can you define YA and middle grades genres like when you have all this data like what can we actually extracted as features that differentiate those. I'm looking at depictions of sexuality across age ranges, so like in the Sweet Valley Twins when they're flirting with boys like how is that sort of portrayed differently than in like Sweet Valley University.

Looking at like the juxtaposition of sex and sexual assault and alcohol, as they often occur in these books and then micro-genres and didacticism. Like how often are these books just like straight up lecturing, you know, like readers, as if it was like a passage out of like Our Bodies, Ourselves. So one of the things I put together last night was looking at a bunch of these kind of series or universe books, versus the books that are recommended by the CCB and the various other bibliographies, and I'm grateful to Sara for sharing these bibliographies and including some older ones to help me put together this corpus. So, we have the Baby-Sitters Club corpus plus about 200 other CCB recommended books. Plus, um, you know, Sweet Valley High, and then a bunch of Star Wars middle grade and YA books.

And as you can see here, this sort of teal in the middle is the recommended books, and it shades into some of these series areas to varying degrees. With the Baby-Sitters Club, not very much. These are the three that are really in the same heart of the Baby-Sitters Club in terms of the top thousand nouns that are used across like all these books together and looking at the different frequencies of these like top thousand nouns which is a pretty decent way to get a sense of what these books are about. So these are these are the three that are the only three that occur kind of like in the heart of like the Baby-Sitters Club land.

If you go up to the Sweet Valley land you have a number of different books about more serious and complicated issues here. And if you head over to Star Wars land, you have you know sort of fantasy things and interestingly, there's a lot of dudes in the lower left corner of this graph. Like the amount of men writing books and the way that they tend to cluster in the bottom left, where there is kind of more overlap with these recommended books I thought it was sort of interesting and worthy of further thought. So yeah, we would love to hear your thoughts and ideas! We're working on new books all the time. YAMS is basically just like a fountain of ideas

and possibilities for thinking more broadly about, you know, modern children's and middle grade and young adult novels and we'd love to hear what you think.

Q&A:

Sara Schwebel: Thank you so much! That was fabulous. We're going to go ahead and open up the chat, so please ask any questions of the panelists by typing directly into the chat and I think that we're also going to launch the poll as well which didn't take off at the beginning. So if you would just take a moment to register, who you are, we would appreciate that too.

So I'll start off just with an observation. I was fascinated Quinn that near last slide about novels that aligns in terms of vocabulary you said the top thousand nouns for Baby-Sitters Club and one of them I think was Grace Lin's book *Year of the Dog*. Yeah, so that really surprised me because I would have thought, I think I would have placed the Baby-Sitters Club at an older reading level. I imagine people wonder, you know, middle schoolers wandering around the mall and browsing Borders or Barnes and Noble and maybe picking up copies. And *Year of the Dog*, I just imagine it's a much younger grade level, so I'm curious if any of your questions have investigated around reading levels and complexity of sentences, complexity of prose, as well as vocabulary.

Quinn Dombrowski:

Yeah, I mean that's something that... There are a lot of questions that are really hard to translate into like computational queries but luckily reading level is really one of the easier ones, or at least like that. I mean I've we've gotten into this a little bit with the YAMS more broadly were like there apparently are like 15 different ways to calculate reading level but we can do all of them, like you know, on 15,000 books in about a day, if we just sort of let the supercomputer at it.

Um, yeah I mean it I guess it doesn't surprise me. We haven't looked at it that much among the Baby-Sitters Club they tend to be like the age ranges are often written on the books and I think it's like 9 to 12 is often mentioned here. But I mean, I was reading them in first grade, I think we all read them at different ages and at varying points but, you know, I think, as much as anything, at least for some people, these, books were less like reading about one's peers and more like reading about the cool older girls and kind of like a primer on how to be a tween for people who were not there yet.

Katia Bowers: So one of the things about the Baby-Sitters Club is that you might think that they're often spending time in malls and doing all this teenager stuff, but a lot of their time is spent babysitting. So there is a lot of vocabulary about children's play and thinking about children and that I think has something to do with this too. Right, it's not quite the vocabulary you would think it is.

Roopika Risam: Can I add something as well? I mean just in terms of, to put on my Education Professor hat on for a second, but that I think when we also think about the measures of reading and things like Lexile scores. It's also worth taking into account that they presume a particular kind of normative reader and normative reading level and when we're actually in, you know, in an environment with lots of multilingual young people, it's a lot more complicated.

Sarah Schwebel: I see a question in the chat.

From: Yuerong Hu to All panelists : Thank you so much for your presentation! Fabulous and enlightening to me! I have a question about how to identify the books beyond the “exemplars.” In addition to text mining and book list cross-references you mentioned, do you think book reviews from readers can be used for extracting some nuanced “truth”? Thanks a lot!

Quinn Dombrowski:

Sure. I mean, there are lots of, you know, projects and ways of getting at reader responses to things. What people actually like got out of these books. I mean, it's challenging with public sites like good reads, you know, because it's hard to like filter by reader because you can get, you know, people our age writing nostalgically about the time they read this book, 30 years ago, and a teenager today being like, yeah I didn't get it. But you can't necessarily like filter out that way but I mean there are lots of things that you can you can draw on besides the books themselves you know with YAMS more broadly, one of the angles that we're looking at, at the Stanford literary lab is looking at, I forget what the modern ones are called but it's like CliffNotes and Spark Notes on what we used to call them like these guides for students who don't want to actually read the book or who wanted help with their assignments. How do those present back the information and things like that?

There's also fanfic, and what directions readers take other characters. There's lots of, like, wonderful Baby-Sitters Club fanfic, you know, especially like playing out kind of like the queer narrative for Kristy, and things like that so, yeah there's there are lots of ways to pick it.

Lee Skallerup Bessette: I think this also goes to that question of, of how we have taken or if we have taken this sort of literature seriously because that goes into what gets digitized and what's gets preserved. So if you're interested in what 80s, you know late 80s early 90s girls, well they probably were making their own zines and there were magazines or publications that were focused towards that would have that sort of reader reaction I mean I can think of Scholastic's right? The Scholastic books would often have blurbs from kids obviously only ones that were positive about books. But these kinds of things existed. But have they been preserved and digitized? And I think that that's a really important question to ask as we go through this project. That you know, pre-internet sure we can go to the fanfic sites, but I wrote fanfic, but it didn't live on the internet because it didn't really exist, we didn't have access to it. It lived on one of my best friend's old Apple Macintosh computer, and she was the one who had it so we all went to her house for sleep-overs so we could type up our fanfic,

You know, I think that there is a lot of renewed interest in girl culture. And how do we then encourage the preservation and possible digitization for these kinds of questions to be asked? Because they're really hard questions to ask about the “original readers” of the book because it was pre-internet. And these things were circulating a lot of times in private and not even semi publicly but very, very privately.

Roopika Risam:

Lee just made me realize that I wrote New Kids on the Block fanfic when I was in third grade, and I just realized that's what it was. We're getting at some of these topics in a panel we're doing for the Children's Literature Association. At least I'm looking at the live journal, the Baby-Sitters Club where you have adults now who read, for the most part, the Baby-Sitters Club books

when they were young, kind of hate-reading in a sense, and so I think really what's key with what we can extract from it is really thinking about what questions are appropriate to what kind of data set? So I don't think we would necessarily get to any kind of meaningful claims to be made about the books themselves, but we can certainly make claims about the way that this weird kind of hate-nostalgia works or kind of people's reflections on their previous understanding and experience the books and how that's changing particularly as they have different frames of reference by being adults.

Maria Cecire:

And I just want to add that one of the great things that about working on children's literature and in particular about working on realist, popular series fiction and about working through a computational mode is that you know both of these things are kind of fringe-y in the world of literary studies. And so, the kind of double fringy-ness of the project, kind of puts us in a position to be really thinking about what are the methods that we use one, the texts that we focus on, who are the readers, who's strategies we center and prioritize. And this kind of asks us to rethink all of that from the beginning, and gives us an opportunity to step back and instead of taking a lot of that for granted, or getting to embroiled in method wars, which some of you may know who I been flaring om literary studies recently, but to just kind of think from a kind of a different place about what is it that we all are doing here and I feel like that's really rewarding to me because that's something that I'm very interested in. You know what is it that we as literary scholars are doing..

Quinn Dombrowski:

And for more adult reader reflection on the series, this summer, there's a book coming out, *We Are the Baby-Sitters Club*, that we have an essay on.

From Patrick Salkeld to All panelists : What types of resources exist for people interested in starting their own projects, especially if they have no experience/prior knowledge? For instance, I am interested in examining Brian Jacques' *Redwall* series or children's/YA books about soccer.

Quinn Dombrowski:

Congratulations on your project. That's exactly who our website is for. That's one of our audiences, Literally we start with like, how do you get these books and scanning and OCR and all of that and then like, you know, stepping through a set of tools and questions and things like that. So if you check out our website and start reading from the beginning, hopefully that can help get you on the right foot.

Lee Skallerup Bessette:

Because part of it was that there aren't many resources out there. Like we're making this up as we go along, because as Maria said it's fringe-y. It's double fringe-y. I'm doing translation which makes a triple fringe-y. We are at the fringes. The great thing is that it's really liberating in a sense because as Maria said, we're making it up as we go along, but we can make it where we want it to be. And so, we have our way of doing it is one way. Read about it. Think about it and then, you know, maybe you come up with a better way. Write about it and share that too. It's all part of this larger conversation that we're trying to engage in around these questions and these issues around YA and computational research methods.

From Jamillah Gabriel to All panelists : Given the lack of diversity in the characters in the series, it would be interesting to read a chapter on race and how it is treated. Is a future chapter such as this under consideration by any of the scholars?

Roopika Risam:

We do not but these are questions that have been coming up. And the reason we don't is actually we were going to do one. And then it was, I don't know May or June. And Maria and I had a conversation about this and we're just like you know what, there's so much like jovial jokey, you know, tongue-in-cheekiness to what we do with us, that it just felt like that was the wrong time to be doing that. Because of course George Floyd had been murdered. Many other black men and women have been murdered and you know that was not the time for the Data-Sitters Club's interpretation of race in the Baby-Sitters Club books.

However, there are a bunch of questions I'm really interested in. One I think it's come up is like the use of food to describe people who aren't white, like, Claudia has almond shaped eyes, Jesse's cocoa skin. The ways that language around race and how it's articulated changes at particular times. These are questions on our list and it's just been a matter of, you know, I think when's a good time to take this on in the style that we that we do, that's not going to do harm.

Lee Skallerup Bessette:

There's also multiple like the question about as, as we go through multiple adaptations as well, right how it has been adapted for Netflix and how diversity has been put into these versions and how it's been illustrated in the graphic novels and how diversity is put in there. In fact one of the slides, you know, slightly makes reference to that whereas the same scene with different with different artists picture it very differently in terms of the background diversity.

So, and again, Maria points out this isn't necessarily new and adaptations these things are modernizations, but it does represent a subtly addressing it as it goes through the series and as it is modernized and contemporized. So we've been interested in, but as Roopsi said, not something that we've found a way to really talk about yet in a way that would be meaningful and also informative in terms of what we do and how we'd like to do it.

Anouk Lang:

Just to jump in to say that our most recent book, which is on GPT 2 which is a model of language that allows you to generate automatic text which Quinn and I did quite recently. So that is it's mostly on automatic text generation, but if you kind of are aware of the conversations about AI and text generation you'll be aware of that like there's a lot of talk about how these models will reproduce the biases and the know quite often horrendous racism evident in these languages. So we kind of gestured towards these debates. Some people at Google just lost their jobs because of this. So we're trying to signal these things, but we are trying to take the right time and with the right knowledge under our belts.