

In Quest of Transition Books

Denis Kotkov¹, Kati Launis², and Mats Neovius¹

¹ Åbo Akademi University, Dept. of Computer Science, Turku, Finland

² University of Eastern Finland, Dept. of Literature, Joensuu, Finland
kotkov.denis.ig@gmail.com, klaunis@utu.fi, mats.neovius@abo.fi

Abstract. Literature read by a person not only reflects, but also affects that person. In fact, certain books (transition books) might trigger this process of becoming interested in grownup's literature and therefore mentally becoming a grownup. In this paper, we detect books that are likely to be transition books or transition book candidates based on a loan dataset provided to us by Vantaa City Library. With four methods applied to this dataset we show what books and why are likely to be the candidates. We found the following candidate books: *Tähtiin kirjoitettu virhe* by John Green, *Punainen kuin veri* by Salla Simukka and *Luukaupunki* by Cassandra Clare. Our findings also indicate a few other books that are less likely, but still good candidates for transition books.

Keywords: Literature, Data Analysis, Growing Up, Transition Books.

1 Introduction

The status of an adult or child gives different rights and responsibilities to their holders, while this status usually depends on the chronological age of a person, which does not always correspond to their mental age. Meanwhile, literature that the person reads both reflects and affects their mental age [2], which suggests that certain books might trigger the process of a child mentally becoming an adult, and mental age then might be assessed based on the literature the person reads. In this paper, the term *transition* refers to an individual becoming interested in literature written and consumed predominantly by grownups. The first book for grownups, after which the reader who used to read books for children starts reading also books for grownups is referred to as *transition book*. The transition process can also be triggered by other factors. However, in this paper, we only focus on books.

In this paper, we look for books that are likely to be transition books (transition book candidates). We do not detect the actual transition books, as this study is just the first step in detecting transition books. We apply four methods to detect and investigate the candidates based on the dataset provided to us by Vantaa City Library. According to our analysis, the following books are the candidates: *Tähtiin kirjoitettu virhe* by John Green, *Punainen kuin veri* by Salla Simukka and *Luukaupunki* by Cassandra Clare.

The paper is organized as follows. We review literature in section 2, present the dataset for our analysis in section 3, detect transition age in section 4, which helps us detect transition candidates in section 6 with methods described in section 5. Finally, we discuss limitations of our study in section 7 and conclude in section 8.

2 Related works

The culture of reading is changing rapidly today. The importance of reading and concerns on declining literature consumption especially among young people are being widely discussed in research and media [6, 1]. A number of studies have been dedicated to (1) the reading culture of young people (both in schools and during the leisure time) [7, 10], (2) the investigation of young adult fiction, the “transition genre” between children’s and grownups’ literature, written for teens but read also by every generation [10], and (3) reaching reluctant young readers, hooking them and luring to read the entire book [12]. However, to the best of our knowledge, there have not been any studies on the process of transitioning from the literature for children to the literature for grownups on the ground of the big data.

The transition process is a specific moment in a young reader’s “reading career”, which may advance or even finish the urge to read. Studying this process is possible on the ground of the library loan data and the computational methods we are utilizing in this article. This combination of big data and computational methods has hitherto not been used in the Finnish literary studies on reading that has earlier used methods such as interviews and queries [3, 9]. To find the “patterns” and rules behind their reading habits, the daily, big and objective loan data calls for totally new ways of studying the reading culture of the youth. It can also help to encourage the youth to read more or suggest suitable books to them.

3 Dataset

Our dataset has been provided to us by Vantaa City Library. The dataset contains 1,550,260 loans performed between January 8, 2016 and December 10, 2017 by 57,108 library users on 182,612 items, where an item corresponds to an object that can be loaned in the library, such as a CD, a DVD or a book. Each loan is a triple consisting of user id, item id and timestamp when the item was returned to the library. Each user is anonymized and described with user id, birthdate and gender. Item information includes an author, title, language, type (such as a CD, DVD, book), age category, metadata (such as release year and keywords). Age category indicates if the item is meant for grownups (83,497 items), children (27,575 items), both (2,387 items) or unknown (69,154 items). The age category has been assigned to items by the library.

First, we removed all the items (55,339 items) that are not books from the dataset, as items of different types, such as a CD or a book have different effects on users. Second, for simplicity and due to the small size of the “both” age category, we labeled all “both” books as books for grownups, as this is the largest age category and therefore is likely to include items that cannot be clearly classified. Third, for analysis purposes, we also removed 13,821 books with the unknown age category. Finally, we merged all the duplicated books in the dataset. The resulting dataset therefore consisted of 1,017,218 loans performed between January 8, 2016 and December 10, 2017 by 54,156 users on 103,644 books (76,815 books for grownups and 26,829 for children).

Figure 1 demonstrates the age distribution of users in the dataset (min: 0, max: 99, mean: 36). The distribution shows the significant number of library users between 7

and 11 years old. The age is calculated as the difference between the birthdate and the end of observation period (December 10, 2017). Gender distribution of the dataset is dominated by females: 35,334 females, 18,747 males, 69 – unknown and 6 others.

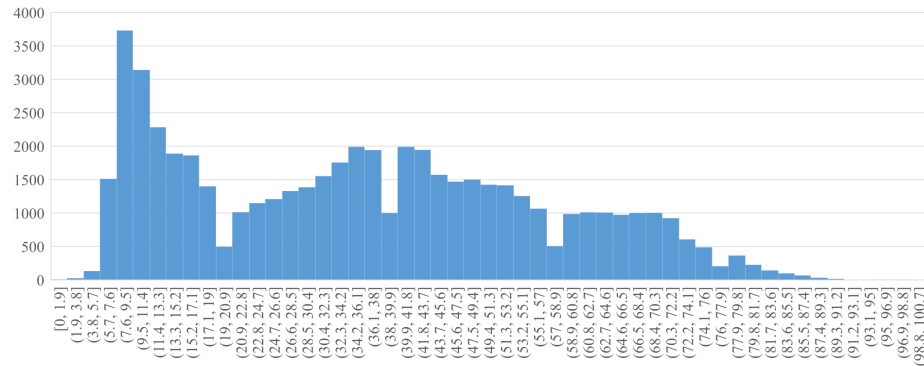


Fig. 1. User age distribution.

4 Detecting transition age

To detect the transition period, we looked at the increase of proportion of grownups' books with respect to the previous year. According to figure 2, the increase from the beginning till age 21 indicates young users growing up and becoming interested in grownups' literature. The following drop of the share around 37 indicates the age period of users loaning children's literature for their kids. Kids then grow up, which is indicated by the rapid increase between 37 and 57. After this age, the share fluctuates, but steadily grows and reaches its maximum.

We selected the transition age based on the chart (figure 2): the first rapid increase indicating the transition period of children becoming interested in grownups' literature. We selected the inclusive age interval between 11 and 17 years, as the increase in the share of grownups' literature was higher than 13% with respect to the previous year. The increase before and after that age was lower than 7%.

5 Methods

To detect transition books, we applied four methods that are based on the frequency of books in different conditions. The conditions filter out more and more books depending on the method and therefore increase the accuracy of the results, but provide weaker signals (lower frequencies). We apply several methods to provide comprehensive data for analysis.

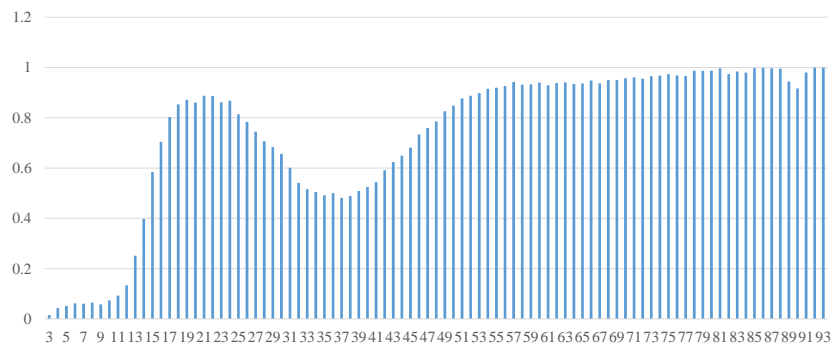


Fig. 2. Share of grownups' literature loaned by users per user age.

- **General popularity.** We order books according to the number of times this book has been loaned by the target user group (users between 11 and 17). If the user loans a book multiple times, we count only one loan. This method detects books mostly loaned by the target group. As many users in the target group are in the transition process, it is likely that the most popular book among them is the transition book.
- **Local popularity.** Similarly to general popularity, we order books according to the number of loans (per user) performed by the target group with the subtracted number of loans performed by other users. The method detects books that are specifically popular in the target group, but not loaned by users outside this group often.
- **Transitional popularity.** Although all the users from the target group are in the transition age, they are not necessarily in the transition phase: some users might already be reading mostly grownup's literature, while some might not have even started reading it. According to figure 3, the majority of users either have already transitioned or have not started yet. This method selects only users whose share of grownups' literature is between 0 and 1 (exclusive interval). We will refer to these users to as *transition users*.
- **Definition-based popularity.** Although books loaned by transition users can trigger the transition process, not all of these books are equally likely to do that. We therefore only considered first grownup's books loaned by transition users that satisfy additional requirements. Users sometimes loan several books at a time. We therefore split user loans into sets of books returned together and selected those transition users whose first set consists of only children's books and at least one consequent set contains at least one grownups' book. In the resulting dataset, we only counted frequency of the books in the first set with at least one grownups' book.

6 Results

Table 1 demonstrates top 10 results generated by each of the four methods applied. The results suggest that the following three books are likely to be transition books as they appeared in each of the four lists:

Table 1. Top 10 results generated by each method. Freq (frequency) corresponds to the number of times the book was loaned based on the method used. The number in brackets (such as (4)) corresponds to number of occurrences. The three books appeared in all lists are highlighted with bold.

General popularity		Local popularity	
Book	Freq	Book	Freq
Tähtiin kirjoitettu virhe / John Green (4)	141	Tähtiin kirjoitettu virhe / John Green (4)	54
Poika raidallisessa pyjamassa / John Boyne (3)	68	Poika raidallisessa pyjamassa / John Boyne (3)	36
Punainen kuin veri / Salla Simukka (4)	54	Vihan liekit / Suzanne Collins (3)	22
Eikä yksikään pelastunut / Agatha Christie (3)	50	Eläinten vallankumous / George Orwell (3)	22
Eläinten vallankumous / George Orwell (3)	49	Punainen kuin veri / Salla Simukka (4)	21
Luukaupunki / Cassandra Clare (4)	45	Maaninkavaara / Miika Nousiainen	16
Hobitti, eli Sinne ja takaisin / J. R. R. Tolkien	45	Luukaupunki / Cassandra Clare (4)	16
Mielensäpahoittaja / Tuomas Kyrö	43	Minne tytöt kadonneet / Leena Lehtolainen	15
Tuntematon sotilas / Väinö Linna	42	Eikä yksikään pelastunut / Agatha Christie (3)	12
Vihan liekit / Suzanne Collins (3)	37	Tummien perhosten koti / Leena Lander	11
Transitional popularity		Definition-based popularity	
Book	Freq	Book	Freq
Tähtiin kirjoitettu virhe / John Green (4)	118	Tähtiin kirjoitettu virhe / John Green (4)	37
Kuroshitsuji - piru hovimestariksi. 2 / Yana Toboso	83	Vihan liekit / Suzanne Collins (3)	16
Emma. 1 / Kaoru Mori	69	Emma. 1 / Kaoru Mori	15
Aron morsiamet. 1 / Kaoru Mori	55	Hobitti eli Sinne ja takaisin / J. R. R. Tolkien	15
Deadman Wonderland. [2] / Jinsei Kataoka, Kazuma Kondou	46	Eikä yksikään pelastunut / Agatha Christie (3)	11
Luukaupunki / Cassandra Clare (4)	43	Luukaupunki / Cassandra Clare (4)	11
Punainen kuin veri / Salla Simukka (4)	43	Eläinten vallankumous / George Orwell (3)	10
Poika raidallisessa pyjamassa / John Boyne (3)	43	Attack on Titan. 1 / Hajime Isayama	10
Mirai nikki. 2 / Sakae Esuno	42	Punainen kuin veri / Salla Simukka (4)	8
Blood+. 01 / Asuka Katsura	41	Kissan tavat tutuksi / Sarah Heath	8

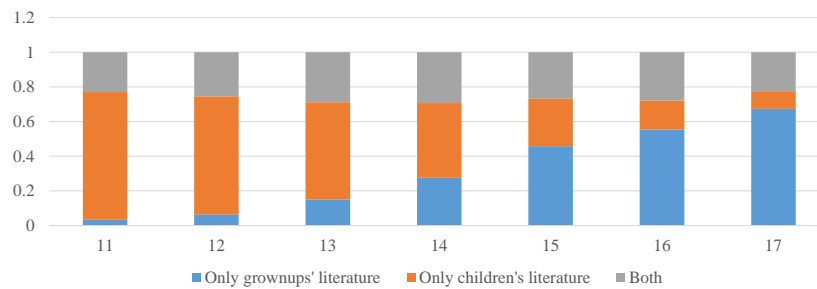


Fig. 3. Share of users who loan particular literature per age.

1. *Tähtiin kirjoitettu virhe* (*The Fault in Our Stars*, 2012) by John Green (translated by Helene Bützow). The book rapidly gained popularity and in 2014 was made into a movie, which also became popular. The young adult novel is about a 16-year-old girl meeting a boy at a cancer patient support group. Cognitive-affective engagement with the fictional characters comes from the narration of the novel [10]. Green's novel uses topics, such as sickness, love, death, sex that are close to the young adults. It has the classic plot of a romance that is ending to death [11] and represents "sick lit", the popular sub-genre of young adult fiction, which focuses on a sick or dying child or adolescent [5]. The young are likely to read Green's novel for enchantment, knowledge or shock [4].
2. *Punainen kuin veri* (*As Red as Blood*, 2013) by Salla Simukka. The Finnish young adult novel tells a fictional thriller story about a 17-year-old girl investigating drug trafficking business in a Tampere, Finland. The thriller is the first part of the popular trilogy whose translation rights have been sold to 52 countries.
3. *Luukaupunki* (*City of Bones*, 2007) by Cassandra Clare (translated by Terhi Leskinen). The urban fantasy novel is set in modern days in New York and tells a story about a 15-year-old girl discovering the world of demon-hunters. *Luukaupunki* is a young adult bestselling novel, which has been made into a movie.

Each of the transition candidates (1) is very popular, (2) belongs to novels for young adults and (3) tells a story about a 15-17-year-old girl. The popularity of these books help them gain visibility among the target audience. The writing style of these novels helps them reach young adults. The age and gender of the main character make it easy for readers to dive into the story (most users are females, section 3).

The following four books are also likely to be the transition books, as they appeared in the three out of four lists:

1. *Eikä yksikään pelastunut* (*And Then There Were None*) by Agatha Christie (translated by Helka Varho). The mystery novel is about eight people arriving to a small isolated island in England.
2. *Eläinten vallankumous* (*Animal Farm*) by George Orwell (translated by Panu Pekkanen). The allegorical novel is dedicated to the story about farm animals trying to create a society of equal and free animals. This classic of juvenile literature is widely read at schools in Finland, and is likely to be loaned for these purposes.

3. *Vihan liekit (Catching Fire)* by Suzanne Collins (translated by Helene Bützow). The dystopian science fiction novel tells a story about 16-year-old girl and boy taking part in a survival game.
4. *Poika raidallisessa pyjamassa (The Boy in the Striped Pyjamas)* by John Boyne (translated by Laura Beck). The novel tells a story about a 9-year-old boy growing up during World War II in Berlin. The book has been released in 2006 followed the movie adaptation in 2008.

The book *Poika raidallisessa pyjamassa* is less likely to be the transition book, as it did not appear in the definition-based popularity list. It suggests that the book is generally popular in the transition age, but it might not trigger the transition process.

The 10 most popular books among transition users (transitional popularity, table 1) include six manga comics: *Kuroshitsuji - piru hovimestariksi. 2*, *Emma. 1*, *Aron morsiamet. 1*, *Deadman Wonderland. [2]*, *Mirai nikki. 2* and *Blood+. 01*. This might be the genre that triggers the transition process for some users. However, these books do not appear in other lists (except for *Emma. 1*), as the popularity is equally spread among these books.

6.1 Transition candidate analysis

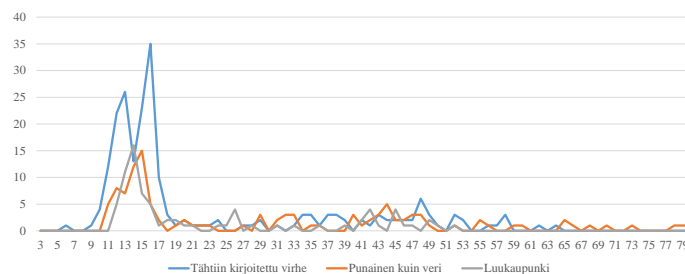


Fig. 4. Book loans per age.

Figure 4 demonstrates book popularity by age. *Tähtiin kirjoitettu virhe* is the most popular book among the three transition candidates. It has been loaned by 215 users (188 females and 27 males) in our dataset. It has two peaks of popularity: loaned by 26 13-year-old users and by 35 16-year-old users. *Punainen kuin veri* has its peak at 15 years (15 users) and overall loaned by 114 users (102 females and 12 males). The peak of *Luukaupunki* is at 14 years (16 users). Overall, this book has been loaned by 79 users (75 females and 4 males).

Each of the candidate books have their peaks within the transition period, which suggests that they are very likely to be the transition books. However, the majority of users loaning these books are females, which indicates that these transition books might be significant in the transition process only for the reference group, the females. Further investigation is needed for a stronger claim.

Transition candidates might be universal, as some discovered books (table 1) appear in other surveys on transition books³, which is supported by the facts that Finnish young readers favor the anglo-american young adult fiction [8], and foreign literature interests young readers [7]. However, in this paper, we can only generalize our findings to Vantaa City Library.

7 Limitations

This paper has a number of limitations. First, the transition process can be triggered by different factors, but in this paper, we only focused on books. Second, we only detected the transition candidates, as further investigation is needed to find out if reading these books actually lead to transitioning from children's to grownups' literature. Third, we only took into account books that users loan in Vantaa City Library. Forth, based on our dataset, we cannot infer if users read or liked books they loaned. Fifth, as we studied Finnish library users, our findings will generalize to mostly these users. Considering individuals, who read books online or buy them in book stores might lead to different results. Finally, the distinction between grownups' and children's literature is not very strict. However, based on our analysis, the labeling used in the dataset seems representative, as it seems to correlate with user age (figure 2).

8 Conclusion

In this paper, we investigated books that might trigger the process of library users who used to consume children's literature loaning their first book tagged as grownups' literature, that is, the transition books. We applied several methods to detect candidate books that are likely to be such transition books. According to our findings and within the dataset used for the analysis, these candidate books are *Tähtiin kirjoitettu virhe* by John Green, *Punainen kuin veri* by Salla Simukka and *Luukaupunki* by Cassandra Clare. These books are likely to be transition books only for females, as they are mostly loaned by females. We also found that the following books are less likely, but can be transition books: *Eikä yksikään pelastunut* by Agatha Christie, *Eläinten vallankumous* by George Orwell, *Vihan liekit* by Suzanne Collins, *Poika raidallisessa pyjamassa* by John Boyne and manga comics, such as *Emma. 1* by Kaoru Mori. The data analysis clearly shows the popularity of the genre of young adult fiction as a "transition genre" among the young readers in the Finnish library. We believe analytical exploration of such transition books is central when the time spent on reading books is challenged by new media.

9 Acknowledgements

The research was performed in the LibDat project supported by the Academy of Finland, grant #309495.

³ <https://www.theguardian.com/books/booksblog/2014/nov/19/what-book-marked-your-transition-from-childhood-to-adulthood>

References

1. 5 faktaa lasten ja nuorten lukemisesta - lukukeskus. <https://lukukeskus.fi/5-faktaa-lukemisesta/>, last accessed 2020/02/09.
2. Billington, J.: Reading between the lines: the benefits of reading for pleasure. Quick Reads, University of Liverpool (2015).
3. Eskola, K.: Suomalaiset kirjanlukijoina. Tammi Helsinki (1979).
4. Felski, R.: Uses of literature. Usa & UK: Blackwell Publishing (2008).
5. Hardstaff, S.: 'with special obligations': Constructions of young adulthood and caregiving in the road to memphis and seventeen against the dealer. In: Ahlbeck, J., Lappalainen, P., Launis, K., Tuohela, K. (eds.) *Childhood, Literature and Science*, pp. 141–152. Routledge (2018).
6. Harju, J.: Osaavatko teinit enää lukea? lukemisen kulttuurit yläkouluikäisten kirjoituksissa. *Kasvatus & Aika* **12**(2), pp. 50–61 (2018).
7. Herkman, J., Vainikka, E.: Lukemisen tavat: lukeminen sosiaalisen median aikakaudella. Tampereen yliopisto (2012).
8. Launis, K., Aino, M.: Mitä tehdä, kun Shakespeare ei vloggaa eikä Waltari twiittaa? Koulu, kirjasto ja nuorten uudistuvat lukemiskulttuurit. Järjestyksiä kirjahyllyssä. Kirjallinen elämä markkinaperustaisessa mediayhteiskunnassa. Toim. Elina Arminen, Anna Logrén Erkki Sevänen (in press). Tampere: Vastapaino (2020).
9. Neovius, M., Launis, K., Nurmi, O.: Exploring library loan data for modelling the reading culture: Project libdat. In: 3rd Conference on Digital Humanities in the Nordic Countries, DHN 2018; Helsinki; Finland; 7 March 2018 through 9 March 2018 (2018).
10. Nikolajeva, M.: Visible, audible and sentient: Cognitive-affective engagement with disability in contemporary young adult fiction. *Childhood, Literature and Science: Fragile Subjects* (2018).
11. Pearce, L., Stacey, J.: *Romance revisited*. NYU Press (1995).
12. Reid, R.: *Reaching Reluctant Young Readers*. Rowman & Littlefield (2017).