

# Audiobook stylistics: Comparing print and audio in the bestselling segment

Karl Berglund, Mats Dahllöf

Karl Berglund, Mads Dahllöf, Uppsala University, Sweden

Peer-Reviewers: Matthew Rubery, Mark Algee-Hewitt

Data Repository: <https://doi.org/10.7910/DVN/MK6SXT>

Article DOI: <https://doi.org/10.22148/001c.29802>

## ABSTRACT

---

The paper explores differences between bestsellers in print and the most popular audiobooks in a subscription-based streaming service for books (“beststreamers”) by means of computational stylistics. The point of departure is the complete set of print bestsellers and digital audiobook beststreamers for the Swedish book market 2015–2019, in total 172 novels. We probed 34 linguistic measures to track differences between subsets at the stylistic level. The results indicate that there are pronounced differences between the formats. Print bestsellers are longer, syntactically more complex and varied, and seem to focus more on depiction. Beststreaming audiobooks, by contrast, are shorter, more straightforwardly written, and appear to highlight plot and dialogue. The results are replicated when the comparison is restricted to crime fiction, the most prominent genre in the commercial top segment. Given these results, it is argued that it is possible to discern a particular audiobook style as one factor affecting book consumption in digital formats, and conversely that the printed format is associated with other stylistic preferences.

---

This article is an investigation of the stylistic differences between works of fiction which are successful in the print format, on the one hand, and those successful as audiobooks, on the other. This is a highly important issue from the point of view of the sociology of literature as audiobooks in streaming services in recent years have become an increasingly prominent format for consumption of fiction. This is having a profound impact on business models, i.e., on how literature is sold and distributed. It will also influence the way in which fiction consumers engage with texts and conceptualize their literature experience. The shift to a situation where the audio format takes center stage is likely to change the norms for how some kinds of literature are written and appreciated. The present study uses the complete set of print bestsellers and digital audiobook beststreamers for the Swedish book market 2015–2019 as its empirical material.

Over the last few decades, book production and distribution have moved towards digital formats and channels. First, digital sales platforms of print books emerged, leading to a revolution in book distribution. Then, literature started to be consumed in digital book formats – as ebooks, obviously, but in the last few years increasingly as digital audiobooks, i.e. as audio recordings of literary text. Today audiobooks are mainly consumed through online streaming platforms, such as Audible and Storytel.<sup>1</sup>

The growing popularity of streamed audiobooks can to a great extent be explained by the fact that the medium is both portable and consumed aurally. With a smartphone and a pair of earphones anyone can listen to an audiobook anywhere, and at any time – also while doing something else. In this respect, audiobook consumption resembles listening to radio and podcasts (whose growing popularity is another significant global trend) more than consumption of print books, ebooks, films, and tv series.<sup>2</sup> This is important, as it affects book consumption behavior. Even though the text of the works is the same, it is a different thing to listen to an audiobook than to read the text version.

The audiobook boom is a global trend in publishing.<sup>3</sup> In Sweden, this trend has coincided with the rapid rise of subscription-based streaming services for books, and a remarkable growth in the consumption of streamed audiobooks.<sup>4</sup> In 2020, a symbolic threshold in the trade was crossed when subscription-based streaming services were reported to constitute a majority of the total number of volumes sold in the market: 57%.<sup>5</sup> And, in contrast to US-based services like Scribd and Kindle Unlimited, book streaming services in Sweden are (almost) all about audiobooks. The by far largest actor, Storytel, holding around 70% of the market share for book streaming services in Sweden, reported that 92% of the consumption on the platform in 2020 was made up of audiobooks.<sup>6</sup>

In Sweden, then, ebooks are a relatively insignificant phenomenon, while streamed audiobooks make up roughly half of the market in terms of volumes.<sup>7</sup> When it comes to revenues, the market share is smaller, because the subscription model returns less revenue to publishers than print copies.<sup>8</sup> But this does not take away the fact that streamed audiobooks have now become one of the major formats for book consumption in Sweden.

This transformation of the publishing industry has led to heated debates. In 2018, a conflict escalated at Sweden's largest publishing house, Bonniers. The CEO claimed that "Nobel Prize winners underperform in digital formats," a statement that was formulated as a problem that needed to be solved. Key employees left because they felt compelled to adjust their editing practices and publishing decisions to better serve the audiobook format. Several Swedish intellectuals emphasized the cultural importance of the print book. Some even mocked the audiobook as being a format "for grown-ups longing for bedtime stories."<sup>9</sup>

In 2021, a similar debate emerged. Some commentators argued that listening to books is not reading and should therefore not be associated with the positive aspects connected to book reading (in terms of reading ability, language development, etc.). Moreover, they stressed that primarily popular genre fiction gets consumed on these platforms.<sup>10</sup> Others claimed, on the contrary, that audiobook consumption does not primarily compete with print reading. Rather, it supports literature in general by stimulating book consumption and leading to a growing book trade.<sup>11</sup>

These debates can be connected to the longer history of audiobooks and opinions about the format in the world of books. As Matthew Rubery has shown, audiobooks have for long been linked to reading disabilities and, in accordance with this, primarily been understood as a second-rate medium for literature, as books for those who are prevented from reading in other ways.<sup>12</sup> Such conceptions are apparently still present in debates. Even so, the commercial impact of streamed audiobooks has started to change this, at least in Sweden, and most definitely among publishers, who have come to realize the potential and importance of the format. Storytel and other streaming services have started to produce their own content directly for the audiobook format. Large publishers such as Norstedts and Bonniers have imprints (e.g., Tiden and Bonnier Bookery) dedicated solely to what in the publishing business is labelled "born-audio" or "audio first" works, i.e., works written directly for the audiobook format. And all publishers in Sweden today are very aware of the fact that audiobooks make up such a large proportion of the market. Books that are commercially important, then, need to perform well as streamed audio, to paraphrase the Bonnier CEO above.

The indignation from authors and publishers in the Swedish audiobook debates seems to be based on an assumption that audiobooks popular in streaming services are less complex, less literary, and more streamlined than popular print books. But can this be measured? What kinds of patterns can be found when we compare books popular in audio and books popular in print on a larger scale?

## **Purpose and Aims**

The purpose of this article is to study the impact of format – print or audio – on consumer choices of literary works by contrasting the most successful titles for each format. To be more specific, the article makes a stylometric comparison of bestsellers in print and “beststreamers,” i.e., the most streamed audiobooks. The concept of the beststreamer also corresponds to a measure of actual book consumption, as it tracks reader engagement in terms of streaming patterns rather than mere sales figures.<sup>13</sup> In this way, we can reach an understanding of the two formats without consulting data which more directly capture consumer behavior. The study covers the complete set of bestsellers and beststreamers for the Swedish book market 2015–2019, i.e., the entire commercial top segment.

The methodical set up is limited to measuring only textual differences. The best-performance data will to some extent reflect text-related preferences, as textual factors will influence which works, series, and authors readers/listeners appreciate. However, their choices also depend on text-external factors and perceptions, relating to genre, performing narrator, availability, marketing, platform design, recommendation systems, and digital subscription models.

Such parameters have been discussed quite a lot in previous studies on audiobooks and streaming services (see below). What we find to be missing is more systematic research on stylistic differences. Our hope is that this study will provide new knowledge on the importance of text-related preferences when choosing which works to consume in the two formats, and how such preferences work in relation to extra-textual factors. Sweden is a case study of particular interest since streamed audiobooks have such a strong commercial position.

The purpose can be broken down into the following more specific research questions:

- Are there textual differences (e.g., regarding style, lexical complexity, lexical richness) between bestsellers and beststreamers?
- If so, would we find similar differences (between bestsellers and beststreamers) if we also restrict the comparison to the most prominent genre in the commercial top segment, i.e., crime fiction.
- How can the findings be understood in relation to literature consumption and the contemporary book trade context?

Digital publishing, digital book formats, digital reading, and digital sales platforms are aspects of the contemporary trade that have been widely discussed in recent works of publishing studies and contemporary book history.<sup>14</sup> The same can to some extent be said about specific digital formats such as ebooks<sup>15</sup> and audiobooks,<sup>16</sup> as well as digital streaming services.<sup>17</sup>

However, previous studies have seldom analyzed literary texts to investigate the influence of digital book formats and changes in the book trade. As Clayton Childress has pointed out, there is often a large gap between scholars of cultural content and scholars of cultural production.<sup>18</sup> Similarly, computational text analysis is an underused method in book history and sociology of literature.<sup>19</sup>

In our comparisons, we treat both print books and audiobooks as *texts*, disregarding the narrator and other aspects of audiobooks *as audio*, but it should be stressed that these aspects of audiobooks are a crucial part of the medium as such.<sup>20</sup> What we are able to track are stylistic differences between the formats, which can then be discussed in relation to various hypotheses concerning how the production and consumption of literature are adjusting to an ecosystem where the audiobook is becoming a major vehicle for literature. In the interpretative parts, we also draw from previous research when it comes to, for example, the role of the narrator,<sup>21</sup> the importance of platform design and recommendation systems,<sup>22</sup> and the characteristics of born-audio works.<sup>23</sup>

## Description of the Corpus

The empirical point of departure for the present study is the complete collection of bestsellers and beststreamers in the Swedish book market 2015–2019.<sup>24</sup> The bestsellers are the novels that have been bestsellers in print in Sweden, according to the Swedish Publishers' Association (SvF), either in hardback or paperback. The beststreamers are the novels that have been streamed the most times among the audiobooks in the Storytel platform.<sup>25</sup> The corpus comprises 172 novels: 83 bestsellers only, 49 beststreamers only, and 40 titles with format-crossover popularity. All in all, the corpus covers the commercial top segment for fiction in its entirety and in all important formats in Sweden for the period studied. Thus, our corpus is not very large, but *complete* given the parameters defining it. In line with Katherine Bode, among others, we hold that the combination of a well-curated dataset and qualitative knowledge of the literary sources involved holds great potential for cultural analytics.<sup>26</sup>

In Sweden (as in most other national book trades), books in the commercial top segment are nowadays published simultaneously as print books, ebooks, and audiobooks, and all versions are unabridged and based on the same text source. All popular titles in the beststreamer corpus are thus unabridged versions of the print original.<sup>27</sup> The texts in our corpus were extracted from e-book editions, published in epub format.<sup>28</sup> They comprise both Swedish originals and translations. Metadata and peritext were manually removed. Since we have cleaned our corpus manually, there is little noise in the text data.

Genres do not distribute evenly over the categories of bestsellers and beststreamers, as can be seen in Table 1. In particular, beststreamers and crossover works are heavily dominated by crime fiction and contain no prestige fiction at all.<sup>29</sup> Similarly, the proportion of translated works is much higher among bestsellers in print than in titles popular in streaming services (see Table 2). A less striking pattern is the distribution of author gender, which is fairly well-balanced in the corpus as a whole, as well as in the three format subsets. However, one can notice a slight male bias in the bestseller subset, and a slight female bias in the beststreamer and crossover subsets.<sup>30</sup>

	<i>Crime fiction</i>	<i>Prestige fiction</i>	<i>Other fiction</i>	<i>Total</i>
<i>Bestsellers</i>	33 (40%)	12 (14%)	38 (46%)	83 (100%)
<i>Beststreamers</i>	45 (92%)	0 (0%)	4 (8%)	49 (100%)
<i>Crossovers</i>	34 (85%)	0 (0%)	6 (15%)	40 (100%)
<i>Total</i>	112 (65%)	12 (7%)	48 (28%)	172 (100%)

Table 1. Genre proportions in the corpus per format category

	<i>Translations</i>	<i>Swedish originals</i>	<i>Total</i>
<i>Bestsellers</i>	37 (45%)	46 (55%)	83 (100%)
<i>Beststreamers</i>	5 (10%)	44 (90%)	49 (100%)
<i>Crossovers</i>	4 (10%)	36 (90%)	40 (100%)
<i>Total</i>	46 (27%)	126 (73%)	172 (100%)

Table 2. Proportions of translations in the corpus per format category

Moreover, there are large differences between the formats on the author level, even when these authors seem to write similar kinds of books. There is no obvious way to explain why some authors of crime fiction and other popular fiction are more successful in print and others in streamed audio.<sup>31</sup> The crossover titles are mainly written by the commercial giants and literary celebrities in the Swedish world of books, e.g., Mari Jungstedt, Lars Kepler, David Lagercrantz, Jens Lapidus, and Camilla Läckberg. The majority of the crossover titles are works of Swedish crime fiction. Many of these have also been successful internationally.

The largest publishing houses have unsurprisingly published the largest number of titles in each of the subsets. The crossover subset is the least diverse in terms of publishing houses, which is related to the fact that the biggest author names tend to be published by the biggest publishers. When bestsellers and beststreamers are compared, however, several examples of publishers with successful books in only or almost only one of the two categories stand out.<sup>32</sup> There is probably a blend of reasons for this result: audiobooks strategies, choice of performing narrators, integration on the Storytel platform, and outcomes of recommendation systems, to name a handful. *Svart stjärna – SIEI* by Jesper Ersgård & Joakim Ersgård is the only born-audio title in the corpus.

Bestsellers, on the one hand, and beststreamers and crossover titles, on the other, are clearly different as regards genre and translation composition. This indicates that consumers choose different kinds of fiction for different formats, and that the audiobook format in particular seems to attract consumption of a certain kind of literature, at least in the commercial top segment of the book trade.

## **Method: A Statistical Approach to Comparative Stylometry**

The point of departure for the stylometric analysis employed here is a range of measurable textual properties such as the fraction of common nouns, adjectives, or verbs among the words. We will call these properties (stylometric) “features.” In the parlance of statistics, they are “variables” associated with literary works. In order not to rely on an assumption that these variables are normally distributed, we make use of non-parametric statistics only. This kind of statistics is based on the fact that such a feature ranks the texts, i.e., puts them in an order according to the value of the feature for the text. This makes it possible to compare categories in terms of how they rank collectively in relation to each other with respect to some feature.

A way of quantifying the dependency between a feature and membership in either of two categories is to use the notion of “common language effect size” (henceforth “CLES”), which was introduced as a straightforward concept for “communicating effect size to audiences untutored in statistics.”<sup>33</sup> The CLES value is the probability that a feature score sampled at random from one category is larger than a score sampled from the other category. Values close to 50% correspond to independence, whereas a considerably higher value indicates a positive correlation between the feature and the first category (and, consequently, that values < 50% capture a positive correlation between the feature and the second category). For instance, as we will see below, if we sample two books, one bestseller and one beststreamer, the probability is roughly 80% that proper nouns are more relatively frequent in the beststreamer than in the bestseller. This example illustrates how the CLES concept in a natural way can be used in the field of stylometry. Its application is simply based on the notion that texts have more or less of various linguistic properties.

We decided to focus on features based on grammatical properties. These are universally valid for all genres, but still capture basic aspects of style and text



structure. More specifically, we used features based on the “Swedish annotation pipeline” (compiled by Robert Östling).<sup>34</sup> In particular, the features derive from its tokenization, part of speech tagging, lemmatization, and dependency parsing. We recorded relative frequencies of parts of speech. The annotation by means of the pipeline gives us 25 different lexical categories (parts of speech). We also investigated basic structural features: total word count, mean paragraph length, and mean dependency depth, for each book. The dependency depth feature is the average number of dependency arcs connecting each word to the dependency root node. This feature quantifies the degree of syntactic embedding. Finally, our analysis involves a number of features capturing lexical diversity. Here, we identify lexical types with lemmas, as assigned by the annotation pipeline. We compute the type-token ratio (TTR) for all words, and for nouns, verbs, and adjectives separately, based on the initial 10,000 words of each book (to adjust for differences in length among individual works). We also examine the proportion of common and rare words.<sup>35</sup> We thus describe each novel by 34 features: 25 features relating to lexical category, 6 relating to lexical diversity, and 3 relating to basic structure. This enables the same number of comparisons between bestsellers and beststreamers.

The CLES score captures how two different categories (e.g., of genres and book formats) position themselves collectively in relation to each other when ranked according to one of these features. Here, we discuss CLES scores larger than 64.5%.<sup>36</sup> This is a somewhat arbitrary and pragmatic threshold, which gave us a reasonable number of quite strongly attested tendencies. As our study is based on a total population survey, the problem of sampling bias does not arise, and conventional significance testing is thus not applicable.<sup>37</sup>

## **Results: Differences in Stylistic Properties**

When we compare the bestsellers (83) and the beststreamers (49) (excluding crossover works belonging to both categories), we find six features correlating positively (above our threshold) with the bestseller category, as shown in Table 3. Table 4 exhibits the features positively associated with the beststreamer category.

<b>Feature</b>	<b>CLES</b>
(mean) dependency depth	73.5
ordinal numeral (ro)	72.4
adjective (jj)	71.2
foreign language word (uo)	65.0
intersentential punctuation (mid)	65.0
document length (word count)	65.0

Table 3. Features giving works in the bestseller category ( $n = 83$ ) a CLES score of  $> 64.5\%$  in relation to beststreamers ( $n = 49$ ): CLES score per feature.

<b>Feature</b>	<b>CLES</b>
proper noun (pm)	79.9
(verb) particle (pl)	69.1
verb (vb)	66.7
interrogative/relative determiner (hd)	65.3
sentence-terminal punctuation (mad)	64.5

Table 4. Features giving works in the beststreamer category a CLES score of  $> 64.5\%$  in relation to bestsellers: CLES score per feature.

Several of these features, e.g., mean dependency depth and intersentential punctuation, suggest that the language of bestsellers tends to be syntactically more involved. As can be expected from this, beststreamers are high in sentence-terminal punctuation, i.e., sentences are on average shorter. Furthermore, bestsellers tend to be longer than beststreamers.

The findings suggest that complex noun phrases are more prominent in the bestsellers, as seen in the higher relative frequencies of ordinal numerals and adjectives. The features which correlate positively with the bestseller category seem to point in the same direction: Higher frequencies of proper nouns, verbs, and verb particles can be expected when noun phrases cover smaller segments of the text. Bestsellers contain more foreign language words, whereas beststreamers tend to include a larger fraction of interrogative/relative determiners. The latter correlation might simply be due to the presence of dialogue rich in questions.

In Figure 1, the works in the corpus are plotted along the two features that most effectively separate bestsellers and beststreamers. We see how print bestsellers rank

higher in mean dependency depth and audiobook beststreamers higher in relative frequency of proper nouns. The figure also shows that bestsellers are more dispersed when it comes to these two features, while the beststreamers and the crossovers cluster more tightly together. This is especially apparent for mean dependency depth, where both the titles with the lowest scores (written by Ulf Lundell, Karin Smirnoff, Linda Olsson) and the ones with the highest (written by Hanya Yanagihara, Vigdis Hjorth, and Kazuo Ishiguro) are bestsellers in print. The most frequent usage of proper nouns is found in beststreamers (written by Swedish crime writers Jonas Moström and author duo Anders Gustavsson and Johan Kant), while the lowest ratios are found in bestsellers, of various kinds – from Nobel-prize laureate Olga Tokarczuk to Alex Schulman, a popular Swedish writer of autofiction. The two red data points furthest out from the audiobook beststreamer cluster (i.e., titles low in proper nouns and fairly high in mean dependency depth) are *Behind Closed Doors* by B.A. Paris and *Educated* by Tara Westover – we will return to these titles.

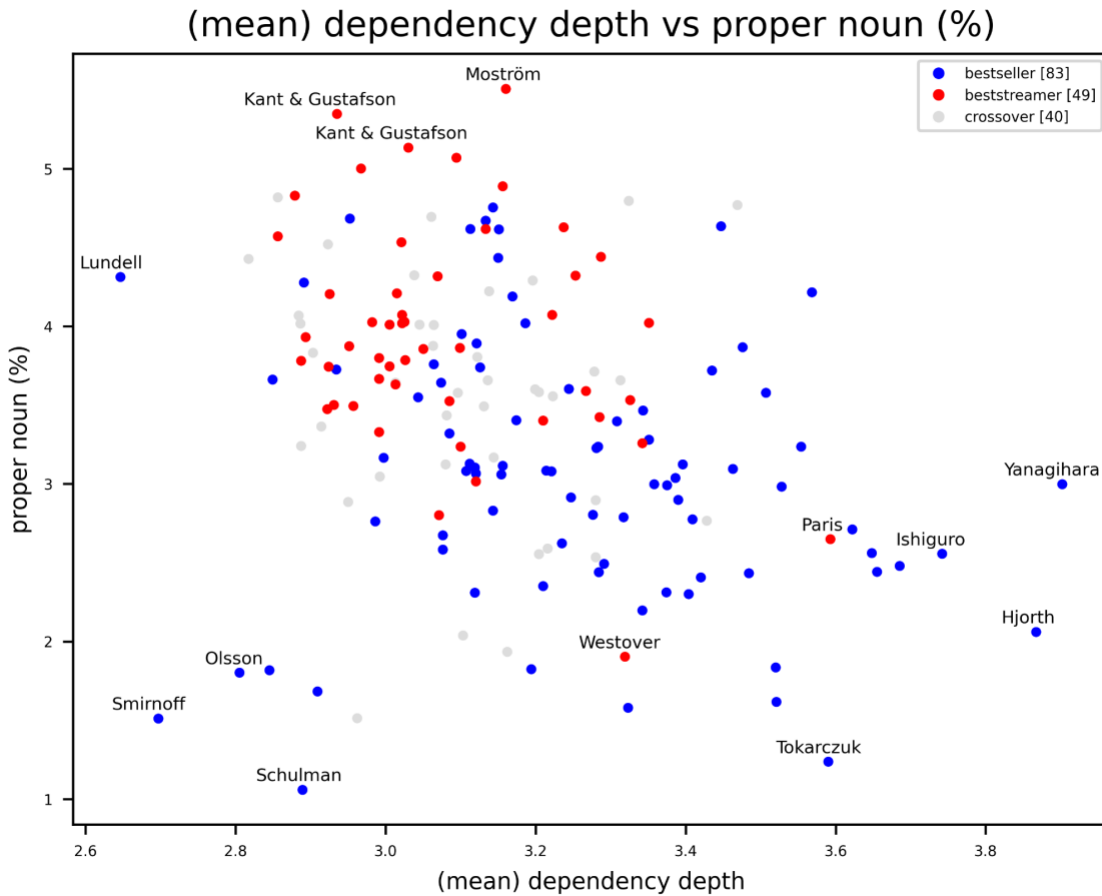


Figure 1. Scatterplot visualization of how the works in the corpus distribute regarding mean dependency depth and relative frequency (%) of proper nouns per popularity format. Selected books labelled with their author(s).

In Figure 2, all works in the corpus have been plotted regarding the relative frequency of sentence-terminal punctuation (STP) and nouns. Again, the bestsellers (red dots) are clearly more tightly clustered around the median point, while the bestsellers (blue dots) are more dispersed. Works along the edges of the distribution are, with one exception, bestsellers. For instance, the works with the longest average sentences (e.g., *Will and Testaments* by Vigdis Hjorth, STP 0.42%, and *A Little Life* by Hanya Yanagihara, STP 0.45%) as well as the ones with the shortest sentences (*Jag for ner till bror* [I travelled down to Bror] by Karin Smirnov, STP 11.1%, and *The Woman in the Window* by A.J. Finn, STP 10.0%) are bestsellers in print. A similar pattern is shown for noun ratios. The highest rates are found among prize-winners (the Pulitzer Prize-awarded novel *All the Light We Cannot See* by Anthony Doerr, and Olga Tokarczuk's *Flights*), and two historical crime novels by the Swedish writer Niklas Natt och Dag. The lowest rates are all found in bestsellers of various genres (e.g., Nobel-prize laureate Kazuo Ishiguro and British romantic fiction writer Lucy Diamond), with the interesting exception of the uttermost outlier, B.A. Paris' *Behind Closed Doors*.

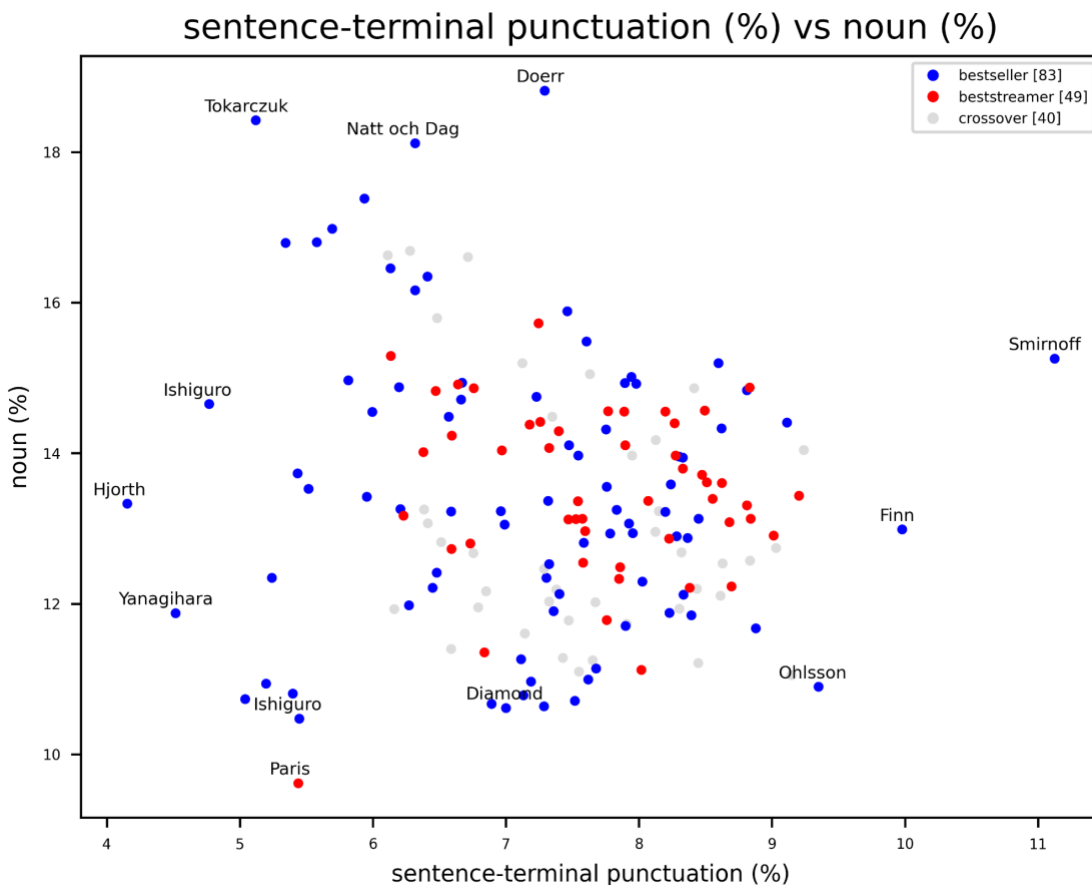


Figure 2. Scatterplot visualization of how the works in the corpus distribute regarding relative frequency (%) of sentence-terminal punctuation and nouns per popularity format. Selected books labelled with their author(s).

When verb and adjective ratios are plotted, a linear dependency is apparent, and the diversity of the bestsellers emerges in a fashion similar to what we saw above (see Figure 3). Regarding these measures, too, streamed audiobooks cluster more tightly together. For adjective ratio, almost all top titles are bestsellers, 25 out of 30, and only one is a beststreamer among the top 30. When it comes to genre, the outliers are more diverse, consisting of Nobel-prize laureate Olga Tokarczuk and Jan Guillou, a celebrity journalist in Sweden who has for decades written popular fiction in various genres. The bottom segment is more mixed concerning formats. To summarize: Popular print titles can be both high and low in adjectives, but popular audiobooks are rarely high in them.

Regarding verb ratio, titles that rank at the top are mostly beststreamers or crossovers, but there are also some exceptions of print bestsellers – most notably *Jag for ner till bror* by Karin Smirnoff (ranked as No. 1, 20.6% verbs) and *Arv och Miljö* by Vigdis Hjorth (No. 4, 20.5% verbs). At the bottom end, however, 22 out of 30 are bestsellers, and only 2 are beststreamers (both are works of crime fiction: *The Hanging Girl* by Jussi Adler-Olsen (No. 17, 16.7% verbs), and *Falleri, fallera, falleralla* by Carin Gerhardsen (No. 30, 17.2% verbs); see Figure 3). Thus, bestsellers in print can be plot-oriented and rich in verbs (as the exceptions show), but beststreaming audiobooks cannot be too low in verbs. It should be noted that B.A. Paris holds an extreme position also regarding these measures. Alongside the Swedish crime writer Emelie Schepp (who is popular both in print and in audio), she is the only author who has written works that are outliers both concerning high rates of verbs and low rates of adjectives.

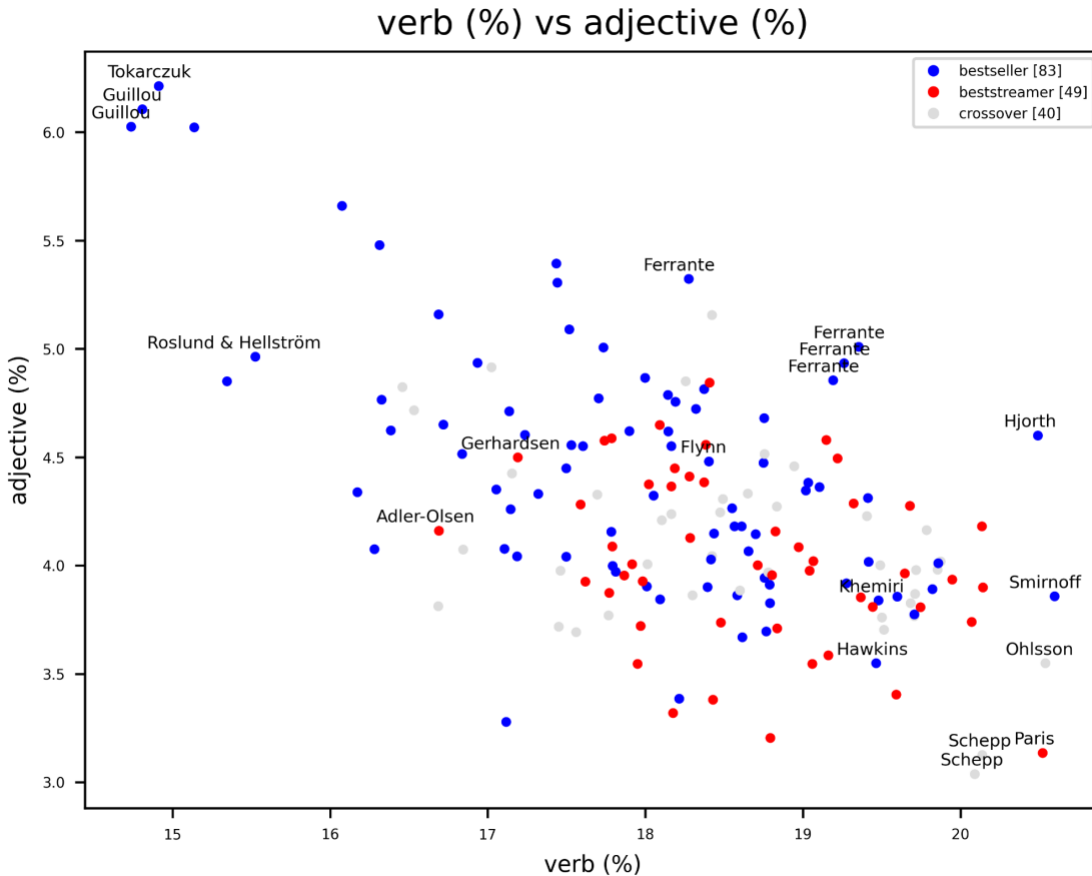


Figure 3. Scatterplot visualization of how the works in the corpus distribute regarding relative frequency (%) of verb and adjective per popularity format. Selected books labelled with their author(s).

The clearest exception from the tighter cluster of beststreamers regarding several key measures highlighted above – *Behind Closed Doors* – is written by the British author B.A. Paris, and originally released in 2016. It is a psychological thriller about Grace Angel, whose seemingly perfect and happy marriage to Jack turns out to be just a facade. In fact, the reader gets to know partially during the narrative, Jack is a total psycho who intends to lock both Grace and her mentally disabled younger sister up in the grand villa that he bought for the family.

The position of this title as a stylistic outlier in the genre is probably due to the fact that the whole narrative is told in first person from the perspective of Grace. It contains lots of her thoughts and inner monologue, and less dialogue, at least when compared to standard contemporary crime fiction prose. A passage typical of the style of the novel is from the opening scene, where Jack and Grace are hosting a dinner for some friends:

I don't know Rufus or Esther. They are new to the area and tonight is the first time we've met, which makes me feel more nervous than I already am. But I can't afford to let Jack down, so I fix a smile on my face, praying that they'll like me. Esther doesn't smile back, so I guess she's reserving judgement. But I can't blame her. Since joining our circle of friends a month ago, I'm sure she's been told over and over again that Grace Angel, wife of brilliant lawyer Jack Angel, is a perfect example of a woman who has it all—the perfect house, the perfect husband, the perfect life. If I were Esther, I'd be wary of me too.<sup>38</sup>

This kind of prose style is uncommon in contemporary crime fiction popular in Sweden, although not unique – in our corpus both *Gone Girl* (originally released 2012) by Gillian Flynn and *The Girl on the Train* (2015) by Paula Hawkins use unreliable first-person narrators (and both lie in the upper half of the corpus regarding dependency depth). What makes *Behind Closed Doors* stand out is the amount of inner monologue; it is its standard form of storytelling. In fact, the snippet above is actually less complex than the means for the novel as a whole. The longest sentence in the quote (46 words) has a dependency depth of 3.96, which is above the mean for the novel (3.59), but at the same time far below its most complex sentences, with dependency depths over 7.

The most complex sentences in the corpus are all found in bestsellers, but also in all genres: prestige fiction (by e.g., Hanya Yanagihara and Jonas Hassen Khemiri), middlebrow literary fiction (by, e.g., Elena Ferrante), crime fiction (by, e.g., Anders Roslund & Börge Hellström), and other popular genre fiction (by, e.g., Jan Guillou). The most complex sentence of them all in terms of dependency depth (13.44) is found in *A Little Life* (2015), an award-winning and bestselling novel by Hanya Yanagihara:

This had begun the second fight, which had started with a confrontation with JB at a cafe near KB's apartment, during which JB had proven maddeningly incapable of apologizing: instead, he talked and talked, about how wonderful the pictures were, and how someday, once he had gotten over whatever issues he had with himself, he'd come to

appreciate them, and how it wasn't even that big a deal, and how he really needed to confront his insecurities, which were groundless anyway, and maybe this would prove helpful in that process, and how everyone except him knew how incredibly great-looking he was, and so shouldn't that tell him something, that maybe—no, *definitely*—he was the one who was wrong about himself, and finally, how the pictures were already done, they were finished, and what did he expect should happen?<sup>39</sup>

In the quote, a scene that occurred earlier and involved two of the protagonists in the novel is being retold from the perspective of one of them. In this context, the text passage can be regarded as an example of a kind of prose style that can work very well in the commercial top segment of print books, but that seems to be less suitable for the streamed audiobook format.

This should not primarily be understood in relation to Yanagihara's novel, however, since the title is one of two in the corpus that have not been made available in the audiobook format, and therefore had no possibility of becoming a beststreamer.<sup>40</sup> The decision by the publisher not to release the novel as an audiobook is hard to uncover beyond speculation. The length of the novel was likely one factor (it is the longest novel in the corpus), as long titles are more expensive to produce. The expected commercial appeal of the title as an audiobook was likely another factor, and here the prose style of the book might have been considered an obstacle for its chances of success as an audiobook.

Rather, the sentence above reflects and exemplifies the general tendency for works ending up as audiobook beststreamers to have less complex sentences. When it is read out loud, this result makes intuitive sense – it is hard to get an oral flow of prose that contains long and complex sentences, such as in the Yanagihara example. Similarly, our results suggest that *Behind Closed Doors* managed to perform well as an audiobook *despite* its more print-oriented style.

The most significant outlier among the beststreamers in terms of another defining feature for the subset, proper nouns, is Tara Westover's *Educated*. It is a literary and self-biographical narrative about growing up in a deeply religious and violent home



in Idaho – in many ways a typical successful middle-brow title, popular among both readers and literary critics. That it ended up in the beststreaming subset is a bit unexpected given the dominance of popular fiction in the format. The explanation lies in how the corpus is defined. As stated, all bestsellers in print in the hardback and paperback format for the period 2015–2019 are included. But paperbacks are often popularized the year after the hardback release, and this was also the case with *Educated*: it did show up in the paperback top list for 2020. Thus, if titles released in 2019 *and* bestsellers in paperback in 2020 had been included in the corpus, *Educated* would have been counted as a crossover title.<sup>41</sup>

### Restricting the Analysis to Crime Fiction

Some of our findings so far might be due to genre-format imbalances, e.g., the overrepresentation of prestigious fiction among bestsellers. Previous studies have found that there are textual differences between popular and prestige fiction, e.g., in the emphasis on plot-related themes in popular fiction and the larger share of descriptions of setting in literary fiction.<sup>42</sup> At the same time, a majority of the crime fiction titles in our corpus are top-listed only in one of the two formats investigated. This suggests that there are differences in consumer preferences also within this one genre. We can thus test the results we obtained by comparing all bestsellers with all beststreamers by restricting the comparison to the crime fiction subsets of the two formats. When we do this, the previous observations about the differences between the two formats are to a great extent replicated (see Tables 5 and 6).

Feature	CLES
document length (word count)	77.8
ordinal numeral (ro)	77.1
(mean) dependency depth	69.2
interrogative/relative pronoun (hp)	65.5
rare words	65.4
intersentential punctuation (mid)	65.4
adjective (jj)	65.0

Table 5. Features giving crime fiction titles in the bestseller category ( $n = 33$ ) a CLES score of  $>64.5\%$  in relation to crime fiction beststreamers ( $n = 45$ ): CLES score per feature.

<b>Feature</b>	<b>CLES</b>
proper noun (pm)	77.0
interrogative/relative determiner (hd)	73.8
verb (vb)	69.6
interjection (in)	65.4

Table 6. Features giving crime fiction titles in the beststreaming category a CLES score of >64.5% in relation to crime fiction bestsellers: CLES score per feature.

The features correlating with beststreamerhood in the crime genre (Table 6) also appeared, with one exception, when we looked at the whole corpus (Table 4). The new feature is the relative frequency of interjections. This part of speech is associated with spoken language. The evidence for a syntactically more involved style in the bestsellers is also replicated (mean dependency depth and intersentential punctuation), as are the signs of complexity in noun phrases (ordinal numerals and adjectives). Crime fiction bestsellers also tend to have higher relative frequency of rare words (somewhat mirroring the corresponding situation with foreign words). The higher share of interrogative/relative pronouns might be another expression of noun-phrase complexity.

On the other hand, crime fiction most popular in the audiobook format to a great extent seems to have the properties we saw previously in beststreamers, in particular high fractions of proper nouns, direct objects, and verbs – all features that indicate plot-driven narratives with plenty of dialogue. The high relative frequencies of interjections – and the dependency relation “discourse” which gives them a syntactic function – also point in the direction of spoken language (see Table 6). A look at rankings on the title/author level for single features show similar patterns. For mean dependency depth, for instance, the top 30 titles are, with one exception, bestsellers (27) or crossover titles (2), even though this list does contain a fair amount of crime fiction (10 titles, 30%).

When we try to control for the possible effect of genre bias by restricting the comparison to crime fiction only, the result remains more or less the same. To some extent, properties of prestige fiction might explain the larger stylistic diversity among popular print books, but they do not explain the stylistic differences as such between print and audio. This adds further support to the hypothesis that audiobook

consumers appear to prefer a different kind of prose style than print-book consumers, at least in the commercial top segment. Crime fiction is more preferred, obviously, but not any kind of crime fiction. What seems to be consumed the most among the most popular audiobooks are the titles in the genre that also suit the format in terms of style.

It should also be noted that we did similar a similar check of the possible impact of translations by restricting the comparison to Swedish originals only. These results to a high degree revealed that the same features correlate positively with the two categories (see the supporting materials) as when we consulted the whole corpus.

## **Discussion and Conclusions**

Based on data covering the Swedish book market during a recent five-year period (2015–2019), this study has shown that there are clearly pronounced stylistic differences between bestsellers in print and beststreaming audiobooks. Print bestsellers tend to be longer, syntactically more complex and varied, and seem to focus more on depiction (visible in, e.g., high adjective rates). By contrast, beststreaming audiobooks are shorter, and appear to highlight plot and dialogue (visible in, e.g., high rates of verbs and proper nouns).

These results are to some extent expected: The genre diversities have been noted before, and several critics and authors have argued that the audiobook boom encourages specific kinds of writing styles.<sup>43</sup> However, the stylistic differences were at the same time surprisingly pronounced. They could also be seen when the comparison was restricted to crime fiction. This lends support to our conclusion that the format factor is the crucial one here, and that the stylistic differences are not actually due to genre differences. This result indicates that there might be reasons to talk about a particular streamed audiobook style, then, as something that differs from a print-book style – at least in the commercial top segment of the book trade.

But what do these stylistic differences tell us? Book consumption is obviously not only driven by textual preferences. At least equally important are extra-textual factors such as material conditions (price, availability etc.), reader habits, perceptions about books, and – in this particular case – perceptions about book

formats and platforms. For our comparison, absolute availability is less important, since all but two of the bestsellers in print are also available as audiobooks. However, far from all titles available in the book trade are available as audiobooks, a bias which affects literary and niche titles more than popular fiction,<sup>44</sup> and which in turn might skew the customer base towards readers more interested in the latter kind of genres. Arguably even more influential factors concerning the customer base are perceptions about audiobooks and subscription-based streaming services.

Unfortunately, we do not know more precisely how the customer base for digital audiobooks is different from the customer base of print books. What we do know is that it is only a minority in Sweden who listen to audiobooks regularly (12%). Print-book consumption is more widespread (34%). Furthermore, people who listen to audiobooks tend to spend more time listening than print readers do reading – they are generally heavier consumers. In terms of reader segments and book consumption, two notable differences between the formats can be noted: the oldest age segment (65–85) consume audiobooks significantly less than they consume print books; and people with a university degree read more print books (and ebooks) than people without a degree, whereas for audiobooks it is the other way around.<sup>45</sup>

Regarding material conditions, it is fairly cheap to subscribe to streaming services for books in Sweden, at least when compared to buying print books. A monthly subscription to Storytel in June 2021 cost 169 SEK (~\$20) a month, which more or less equals buying one hardback or three paperbacks. However, these subscription models are mainly intended for people who own and regularly use a smartphone (which is the device used to stream audiobooks). A reasonable conjecture is that technical obstacles and habits might be the reason for the underrepresentation of the oldest age group, while a combination of factors explain the format differences regarding levels of education (e.g., perceptions about the audiobook format, status, appreciation of the bookish qualities of print books, available titles, reading abilities, audiobook listening possible to combine with work).

Also, the platform itself, with its design, curated categories and list, and underlying recommendation systems, guides consumption behavior. (As do traditional brick-and-mortar bookstores, of course, albeit in different ways.) As highlighted in previous studies, such platforms tend to favor already-successful and well-known

titles, as it suggests similar titles to the ones just finished or continued reading in an ongoing series.<sup>46</sup> At the same time, the lower thresholds for trying out books in subscription services might spur consumers to give new authors a chance. In line with this are the examples of titles from small publishers and self-publishers that have paved their way to the streaming top lists (NoNa, Jentas SE, Bars & Tone). In print and in crossover, all books that sell well come from established publishing houses. Thus, where it is really hard for small actors to become visible in print sales channels, streaming service platforms might provide new possibilities.

Yet another factor is the role of the performing narrators, whose importance for the audiobook format should not be underestimated. A popular narrator can boost consumption for a title, and conversely: certain narrators might be avoided by some consumers.<sup>47</sup>

The results from this study, accordingly, do not show that audiobook consumers consciously choose books of a certain style when they listen to books in the Storytel app. Rather, they show that books of a certain style *are chosen* to a greater extent, and this due to a multitude of reasons. Streamed audiobooks are chosen partly because their style might be a good fit for the audiobook format, but also because of presumptions about their style being a good fit for the audiobook format, as well as because of all other possible aspects at play as per the above.

Why *Behind Closed Doors*, the most distinct outlier among the beststreamers in terms of style, managed to perform well in the audiobook format is probably due to a mixture of reasons, some likely to be extra-textual. But what our results suggest is that the title managed to perform well as an audiobook *despite* its more print-oriented style. The performing narrator of the audiobook version (Kerstin Andersson) is a possible contributing factor, as she is frequently contracted and popular on the Storytel service.<sup>48</sup> That the title was released by the small actor Jentas SE is an important factor in why it didn't manage to become a bestseller in print, and points to the fact that the glass ceiling to the commercial top segment for smaller publishers might be easier to break in streaming services than in traditional book retailing.

As streamed audiobooks are burgeoning in commercial impact, publishers are increasingly observing changes in consumption behavior related to this growing

impact. When they see that certain genres and authors are performing better in the audiobook format, they will draw conclusions from this, based on several factors, one of which is the style of writing.

Up until recently, audiobooks have been understood as a remediation of print books. In 2016, Iben Have and Birgitte Stougaard Pedersen defined the audiobook “as a sound recording of a book that is performed by a professional narrator (often an actor) or the author.” They add: “The audiobook is, as the name indicates, closely related to its written source, the printed book.”<sup>49</sup> More recently, and in the context of the emerging born-audio publishing, this definition has been rephrased, even questioned, by the author themselves and others. In a 2020 paper Stougaard Pedersen argues that “when sound stories are no longer strongly connected to [print] books as cultural objects, they become independent cultural products.”<sup>50</sup>

When seen in relation to the results in this study, however, this discussion can be problematized. In our corpus, born-audio works play a very marginal role. There is only one such title (the Storytel Original title *Svart stjärna – SIEI*). Apart from being born audio, this title is in no way an outlier among the beststreamers: it is popular fiction, written in the stylistic middle segment of most features probed.<sup>51</sup> In the commercial top segment of the book trade, then, there is still a firm connection between print books (and ebooks), and audiobooks – they depart from the same text source, and they are generally released simultaneously. The possibility of switching between formats in streaming service platforms like Storytel – also during consumption, thus moving from minutes listened in the audiobook version to pages read in the ebook ditto, or the other way around – further emphasizes that it is different media *versions* of the same *work* that are being consumed, rather than two separate works.<sup>52</sup>

Yet, the differences in style between popular print books and popular audiobooks suggest that the audiobook format has a deeper impact than that of merely being a matter of handy remediation of print books. If the audiobook was previously regarded as an audio version of the print original, the demands of the format are by now – because of its impact commercially and on actual consumption habits – likely to affect the whole line of text production, including editorial practices and decisions at publishing houses.

This development is confirmed by audiobook authors. Audio-first writer Daniel Åberg claims that audiobook readers have less patience than other readers. He consequently tries to “keep the reader satisfied all the time.” He mentions cutting unnecessary digressions, cutting depictions of milieu, and adjusting dialogue with the reading-out-loud factor constantly in mind.<sup>53</sup> The Australian Audible Original writer Ceridwen Dovey describes the audiobook as “a brutal form regarding lack of attention,” and that it is a format that emphasizes storyline and plot above everything else.<sup>54</sup>

In a similar fashion, Anna Öqvist Ragnar at Storytel states the following on what they want from Storytel Original authors:

Few characters, plenty of dialogue, a forward-leaning storytelling [...]. Shorter and simpler depictions of things, no elaborating [...]. Many external threats, stories that are pushed forward due to events in the plot, not due to what characters feel [...]. Inner monologues and the like we are not particularly interested in. The simpler, and the more eventful, the better.<sup>55</sup>

These voices are anecdotal, but the observations they make seem to be supported by the findings of the present study. What they are describing as the ideal for born audio clearly corresponds to the stylistic tendencies we have seen in beststreaming audiobooks. Thus, consumption patterns in the subscription-based streaming services already seem to favor popular fiction written in this way.

In the future, authors in the commercial top segment will probably start to adapt their writing to what we here call audiobook style: Thus, not primarily born audio, separated from the print world, but audio-adjusted popular fiction, written and edited with the audiobook format in mind, but published simultaneously as print, ebook, and audiobook.

How influential such audio-adjusting will become will probably depend mostly on material conditions. In book markets where the audiobook format is commercially significant, as in the Swedish case, it will likely be of great importance, at least for

popular fiction. Whether similar changes in prose style will start to affect also more prestigious and literary fiction is yet to be seen, but one should not underestimate the power of the reading masses.

As several book historians have pointed out, the novel as we know it (a printed entity that contains one fictional story of a certain length) was not only popularized by the modern book trade, but also shaped by it; it both became and was constructed to be the best-fitted commercial product in the trade.<sup>56</sup> Since the end of the 19<sup>th</sup> century and until very recently, the single-volume printed novel has been the foremost format for production and consumption of literature. The ongoing structural changes in the world of books – which are taking place simultaneously on a distributional level (the rise of digital subscription-based streaming services) and a format level (the rise of audiobooks) – will have an impact on how fiction is written. From a historical perspective this is a rather natural development. Literature is shaped by how it is consumed.

## Appendix A. Details on the corpus composition

	<i>Male authors</i>	<i>Female authors</i>	<i>Gender-mixed author constellations</i>	<i>Total</i>
<i>Bestsellers</i>	43 (52%)	39 (47%)	1 (1%)	83 (100%)
<i>Beststreamers</i>	21 (43%)	26 (53%)	2 (4%)	49 (100%)
<i>Crossovers</i>	14 (35%)	23 (57%)	3 (8%)	40 (100%)
<i>Total</i>	78 (45%)	88 (51%)	6 (4%)	172 (100%)

Table A1. Genre proportions in the corpus by author gender

	<i>Swedish originals</i>	<i>Translations</i>	<i>Total</i>
<i>Crime fiction</i>	98 (88%)	14 (12%)	83 (100%)
<i>Prestige fiction</i>	6 (50%)	6 (50%)	40 (100%)
<i>Other fiction</i>	22 (46%)	26 (54%)	49 (100%)
<i>Total</i>	126 (73%)	46 (27%)	172 (100%)

Table A2. Genre proportions in the corpus per translation



	<i>Swedish originals</i>	<i>Translations</i>	<i>Total</i>
<i>Male</i>	60 (77%)	18 (23%)	83 (100%)
<i>Female</i>	60 (68%)	28 (32%)	49 (100%)
<i>Gender-mixed</i>	6 (100%)	0 (0%)	40 (100%)
<i>Total</i>	126 (73%)	46 (27%)	172 (100%)

Table A3. Author gender in the corpus by translation vs. original language

	<i>Crime fiction</i>	<i>Prestige fiction</i>	<i>Other fiction</i>	<i>Total</i>
<i>Male</i>	51 (65%)	5 (6%)	22 (28%)	78 (100%)
<i>Female</i>	55 (63%)	7 (8%)	26 (29%)	88 (100%)
<i>Gender-mixed</i>	6 (100%)	0 (0%)	0 (0%)	6 (100%)
<i>Total</i>	126 (65%)	12 (7%)	46 (28%)	172 (100%)

Table A4. Author gender in the corpus per genre

## References

1 For summarizing accounts of this development, see e.g. Ted Striphas, *The Late Age of Print: The Late Age of Print: Everyday Book Culture from Consumerism to Control* (New York: Columbia University Press, 2009); Simone Murray, *The Digital Literary Sphere: Reading, Writing, and Selling Books in the Internet Era* (Baltimore: Johns Hopkins University Press, 2018); John B. Thompson, *Book Wars: The Digital Revolution in Publishing* (Cambridge: Polity, 2021).

2 Cf. Lutz Koepnick, “Reading on the Move”, *PMLA* 128 (2013:1); Iben Have, and Birgitte Stougaard Pedersen, *Digital Audiobooks: New Media, Users and Experiences* (New York: Routledge, 2016).

3 Thompson, *Book Wars*: 352–368.

4 For a summary of this development as well as a proposed model for how streamed audiobooks work in the Nordic countries, see Karl Berglund and Sara Tanderup Linkis, “Modelling Subscription-Based Streaming Services for Books” (submitted for publication).

5 Erik Wikberg, *Bokförsäljningsstatistiken: Helåret 2020* (Stockholm: Svenska Förläggareföreningen, 2021), 23–24.

6 Karl Berglund, “Introducing the Beststreamer: Mapping Nuances in Digital Book Consumption at Scale”, *Publishing Research Quarterly* 37:2 (2021): 135–151. <https://doi.org/10.1007/s12109-021-09801-0>.

7 It should be noted that although “volumes” is the term used in the official Swedish book statistics, it is indeed a problematic term to use in relation to subscription-based streaming services, since they do not sell separate volumes but a subscription that grants access to consumption of lots of volumes. What is measured as volumes from the streaming services in the book statistics, then, is titles that have been consumed in their entirety on these platforms. For a more thorough discussion on these matters, see Berglund, “Introducing the Beststreamer”.

8 See Wikberg, *Bokförsäljningsstatistiken*: 17–23; cf. Berglund and Tanderup Linkis, “Modelling Subscription-Based Streaming Services for Books”.

9 Sverker Lenas, and Georg Cederskog, “Konflikt om ljudböcker på Bonniers: ‘Nobelpristagare underpresterar digitalt’”, *Dagens Nyheter*, March 22, 2018 [2018a]; Sverker Lenas, and Georg Cederskog, “Så tycker författarna om ljudboken”, *Dagens Nyheter*, March 23, 2018 [2018b].

10 See e.g. Sven Anders Johansson, “Streamingtjänsterna är dåliga nyheter för litteraturen”, *Aftonbladet*, February 17, 2021; Ola Nilsson, “Lyssna på ljudböcker är inte att läsa”, *Aftonbladet*, March 6, 2021.

11 See e.g. Erik Wikberg, “Djupt problematiskt att en litteraturprofessor sprider falska påståenden om bokbranschen”, *Dagens Nyheter*, February 26, 2021; Alexandra Borg, “Ljudbokens fiender är besatta av status”, *Svenska Dagbladet*, March 10, 2021.

12 Matthew Rubery, *The Untold Story of the Talking Book* (Harvard University Press: Cambridge, MA, 2016).

13 See Berglund, “Introducing the Beststreamer”.

14 See e.g. Striphas, *Algorithmic Culture*; Adriaan van der Weel, *Changing Our Textual Minds: Towards a Digital Order of Knowledge* (Manchester: Manchester University Press, 2011); Padmini Ray Murray, and Claire Squires, “The Digital Publishing Communication Circuit”, *Book 2.0* 3, no. 1 (2013); Matthew Kirschenbaum, and Sarah Werner, “Digital Scholarship and Digital Studies: The State of the Discipline”, *Book History* 17 (2014); Murray, *The Digital Literary Sphere*; Simone Murray, “Secret Agents: Algorithmic Culture, Goodreads and Datafication of the Contemporary Book World”, *European Journal of Cultural Studies* 22 (2019); Ann Steiner, “The Global Book: Micropublishing, Conglomerate Production, and Digital Market Structures”, *Publishing Research Quarterly* 34, no. 1 (2018); Rachel Noorda, and Stevie Marsden, “Twenty-First Century Book Studies: The State of the Discipline”, *Book History* 22 (2019); Thompson, *Book Wars*.

15 See e.g. Simon Rowberry, “Ebookness”, *Convergence* 23, no. 3 (2015); Birgitta Wallin, *Selling and Lending E-Books: Changes in the Swedish Literary Field* (Borås: Högskolan i Borås, 2019); Ann Steiner, “Selling Books and Digital Files: A Comparative Study of the Sales of Books and E-books in Sweden”, *Northern Lights* 13, no. 1 (2015).

16 See e.g. Have and Stougaard Pedersen, *Digital Audiobooks*; Iben Have, and Mille Raaby Jensen, “Audio-bingeing”, *Passage* 83 (Summer 2020); Sara Tanderup Linkis, and Julia Pennlert, “Episodic Listening: Analyzing the Content and Usage of Born-Audio Serial Narratives”, *Journal of Electronic Publishing* 23, no. 1 (2020).

17 See e.g. Terje Colbjørnsen, “The Streaming Network: Conceptualizing Distribution Economy, Technology, and Power in Streaming Media Services”, *Convergence* (October 2020); Karl Berglund, and Ann Steiner, “Is Backlist the New Frontlist? Large-scale Data Analysis of Bestseller Book Consumption in Streaming Services”, *LOGOS: Journal of the World Publishing Community* 32, no. 1 (2021): 7–24.

18 Clayton Childress, *Under the Cover: The Creation, Production, and Reception of a Novel* (Princeton: Princeton University Press, 2017).

19 Cf. Berglund, “Introducing the Beststreamer”.

20 For work focusing on audiobooks as sound and audio, see Lucy Bednar, “Audiobooks and the Reassertion of Orality: Walter J. Ong and Others Revisited”, *CEA Critic* 73 (2010:1); Matthew Rubery (ed.), *Audiobooks, Literature and Sound Culture* (New York: Routledge, 2012); Have and Stougaard Pedersen, *Digital Audiobooks*.

21 Cf. Rubery (ed.), *Audiobooks*; Have and Stougaard Pedersen, *Digital Audiobooks*.

22 Cf. Murray, “Secret Agents”; Berglund and Steiner, “Is Backlist the New Frontlist?”.

23 Cf. Have and Raaby Jensen, “Audio-bingeing”; Tanderup Linkis and Pennlert, “Episodic Listening”.

24 Replication data (comprising code, metadata, and frequency data), along with some supplementary material are found in the Cultural Analytics Dataverse (<https://doi.org/10.7910/DVN/MK6SXT>) and at URL <https://github.com/mdahllof/audiobookstyle>. Unfortunately, we cannot share the our corpus for copyright reasons.

25 Cf. Berglund, “Introducing the Beststreamer”.

26 Katherine Bode, *Reading by Numbers: Recalibrating the Literary Field* (London: Anthem Press, 2012), 12–25; Katherine Bode, *A World of Fiction: Digital Collections and the Future of Literary History* (Ann Arbor: University of Michigan Press, 2018), 17–35.

27 Abridged audiobooks are rare in the Swedish book trade. When original texts have been shortened for the audiobooks format (which is most common for non-fiction and lengthy classics), it is always clearly stated in the paratext. For standard book releases of new titles (as the ones in this corpus), it is commonly ensured in a note after the story ends that the audiobook is based on a verbatim transcription of the print original. Storytel also provides the possibility of switching between the ebook and the audiobook version, which ensures that it is the same text source used. For a description of how the similar state of affairs has become standardized also in the US book trade, see Thompson, *Book Wars*: 364–368.

28 There is one exception to this rule: The born-audio title *Svart stjärna – SIEI* (2016), by Jesper Ersgård and Joakim Ersgård. Since it is a born-audio title produced by the streaming service Storytel, it has never been published as either a print book or an eBook. There is however an ebook version of the title in the Storytel platform. We got access to this title as an epub file directly from Storytel (by email), which we then cleaned from metadata and peritext in accordance with all other titles in our corpus. It should be noted that the title was originally released as a series season, consisting of ten episodes, and that only the first episode made it to the charts.

29 In this study, the categorisation of “crime fiction” follows the bibliography of Swedish crime fiction compiled by the Swedish Crime Fiction Academy. (For further details, see Karl Berglund, *Deckarboomen under lupp: Statistiska perspektiv på svensk kriminallitteratur 1977–2010* (Uppsala: Uppsala University, 2012), 13–17.) Counted as “prestige fiction” are all titles that have received or been nominated for a major literary prize, which have been operationalised as the Nobel Prize in literature, the Pulitzer Prize, the Man Booker Prize, and the two most important literary prizes in a Swedish context: *Augustpriset*, and *Nordiska rådets litteraturpris*. It should be noted that this is an operationalisation of literary prestige that misses some titles that by most critics probably would be regarded as prestige or literary fiction. The label “other fiction” thus applies to all titles not belonging to the two specific

categories. It covers popular fiction in genres other than crime fiction (e.g., romantic comedies, fantasy) as well as middlebrow literary fiction.

30 Since the bias in author gender is minor, it is presented as supporting data and not in the main text. The same goes with the tables showing the relations among genre, translation, and author gender. In short, author gender is fairly well-balanced in relation to both genre and translation. For genres, crime fiction is much more dominated by Swedish originals (88%) than prestige fiction (50%) and other fiction (54%). For further details, see Appendix A.

31 Concerning crime fiction, for instance, Susan Casserfeldt, Mats Ahlstedt and Dag Öhrlund are beststreamers only, where Stefan Ahnhem and Anders Roslund with co-writers are bestsellers only. All titles by these authors are available for streaming in the major streaming services in Sweden, as well as for purchase in the print and eBook format.

32 For instance, all 12 titles from Piratförlaget, a mid-sized publishing house focused on popular fiction and crime fiction, are only bestsellers in print, whereas all 10 titles by Bokfabriken and almost all titles by Lind & Co (13 out of 15) are beststreamers only.

33 K. O. McGraw and S. P. Wong, “A common language effect size statistic”, *Psychological Bulletin* 111, no. 2 (1992): 361–365.

34 Robert Östling, “Part of Speech Tagging: Shallow or Deep Learning?”, *North European Journal of Language Technology* 5 (2018). Resources found at <https://github.com/robertostling/efselab>.

35 The concepts of “common” and “rare” words were operationalized in the following way: We defined “common” as those which are found in at least two occurrences in at least 95% of the works. (The number of actual instances per work will then be much larger.) We also compiled a dictionary of “non-rare” words defined as those which are found in at least two occurrences in at least 5% of the works. (They will consequently include the category of “common” words.). This allowed us to compute the fraction of “rare” words, i.e., those belonging to lemmas not found in the non-rare dictionary.

36 Values lower than 35.5% are equally significant, but we find it easier to discuss these in their “positive” counterparts in relation to the inverse comparison. So, instead of saying that bestsellers outrank beststreamers as regards relative frequency of proper nouns in 20.1% of the pairwise comparisons, we report the factually equivalent percentage ( $79.9\% = 100\% - 20.1\%$ ) based on comparing beststreamers with bestsellers.

37 See Norbert Hirschauer, Sven Grüner, Oliver Mußhoff, Claudia Becker, and Antje Jantsch, “Can p-values be meaningfully interpreted without random sampling?”, *Statistics Surveys*, Vol. 14 (2020), 71–91.

38 B.A. Paris, *Behind Closed Doors* (New York: St. Martin’s Press, 2016). The Swedish translation, which we have used in our experiment, reads: “Jag känner varken Rufus eller Esther. De är nya i området och vi träffas för första gången i kväll, och det gör mig ännu mer nervös än jag redan är. Men jag har inte råd att svika Jack, så jag klistrar fast ett leende i ansiktet och hoppas att de ska tycka om mig. Esther ler inte tillbaka, så hon väntar väl med att bestämma sig. Jag kan inte klandra henne. Sedan hon kom in i bekantskapskretsen för en månad sedan har hon säkert fått höra om och om igen att Grace Angel, hustrun till stjärnadvokaten Jack Angel, är ett perfekt exempel på en kvinna som har allt – det perfekta huset, den perfekta maken, det perfekta livet. Om jag vore Esther skulle jag också akta mig för mig.” (B.A. Paris, *Bakom stängda dörrar*, trans. Ylva Spångberg (Frederiksberg: Jentas SE, 2017): 7–8.)

39 Hanya Yanagihara, *A Little Life* (London: Picador, 2016 [orig. 2015]): 166. The Swedish translation, which we have used in our experiment, reads: “Detta var inledningen till det andra grälet, som börjat med en konfrontation med JB på ett kafé nära JB:s lägenhet, under vilken JB visat sig outhärdligt oförmögen att be om ursäkt: i stället hade han pratat och pratat om hur underbara bilderna var och sagt att Jude en dag, när han väl kommit över de problem han hade med sig själv, skulle uppskatta dem, att det inte ens var någon stor sak, att han verkligen måste ta itu med sin osäkerhet, som i vilket fall var grundlös, och kanske skulle det här bli en hjälp i den processen, att alla utom han förstod hur otroligt snygg han var, så borde inte det säga honom något, att det kanske – nej, definitivt – var han som hade fel om sig själv, och till slut att bilderna redan fanns, de var redan gjorda, och vad ville han?” (Hanya Yanagihara, *Ett litet liv*, transl. Niclas Nilsson (Stockholm: Bonniers, 2016): 167.)

40 The other title not available in the audiobook format is Anthony Doerr’s *All the Light We Cannot See* (2014, in Swedish 2015).

41 The only other title that is affected in the same way regarding our corpus is the crime fiction novel *Victim 2117* by the Danish author Jussi Adler-Olsen. If the numbers for paperback bestsellers for 2020 had been included, also this title would have been counted as a crossover and not as a beststreamer.

42 See Karl Berglund, Mats Dahllöf, and Jerry Määttä, “Apples and Oranges? Large-Scale Thematic Comparisons of Contemporary Swedish Popular and Literary Fiction”, *Samlaren* 140 (2019); Andreas van Cranenburgh, Karina van Dalen-Oskam, and Boris van Zundert, “Vector Space Explorations of Literary Language”, *Lang Resources & Evaluation* 53 (2019).

43 On the genre distribution, see Berglund, “Introducing the Beststreamer”. On audiobooks encouraging specific kinds of writing styles, see Lenas and Cederskog, “Konflikt om ljudböcker”; Daniel Åberg, “At skrive til øret: Betragtninger fra en Storytel Originals-forfatter”, *Passage* 83 (2020): 74–77.

44 See Jerry Määttä, “Kvalitetslitteraturen i luren: Utbudet av Nobelpristagare och Augustnominerad skönlitteratur som strömmade svenska ljudböcker”, in *Från Strindberg till Storytel – korskopplingar mellan ljud och litteratur*, eds. Julia Pennlert and Lars Ilshammar (Göteborg: Daidalos, 2021): 295–326.

45 *Mediebarometern 2020*, ed. Jonas Ohlsson (Gothenburg: Nordicom, 2021): 78–79.

46 See e.g. Ann Steiner, “The Global Book”; Murray, “Secret Agents”; Berglund and Steiner, “Is Backlist the New Frontlist?”; Thompson, *Book Wars*: 176–194.

47 An example of the role of narrators concerns the beststreamers from small publishers and self-publishers. These titles are all published as audiobooks by specific audiobook publishers, for instance Word Audio Publishing, i.e. actors specializing in audiobooks rights and publishing. This model enables titles from indie publishers access to the most popular narrators. For instance, the originally self-published author Susan Casserfelt has had all of her crime novels narrated by Katarina Ewerlöf, the most popular narrator by far on the Storytel platform in the commercial top segment. (See Karl Berglund, “Strömmade bästsäljare. Litteraturkonsumtion i digitala prenumerationstjänster utifrån Storytels användardata”, in *Från Strindberg till Storytel – korskopplingar mellan ljud och litteratur*, Julia Pennlert and Lars Ilshamma, eds. (Göteborg: Daidalos, 2021): 351.

48 Cf. Berglund, “Strömmade bästsäljare”: 351.

49 Have and Stougaard Pedersen, *Digital Audiobooks*: 4.

50 Stougaard Pedersen, “At skrive gennem lyd”, *Passage* 83 (Summer 2020): 86, translation by first author. The original Danish: “Når lydfortællinger ikke længere er tæt knyttet til bogen som kulturelt objekt, bliver de til et selvstændigt kulturprodukt”. See also Tanderup Linkis and Pennlert, “Episodic Listening” for a similar argumentation.

51 The exception is, of course, its short length, due to its being only the first episode in a season of ten. If the length is multiplied by 10, however, *Svart stjärna* has a total word length in the upper part of the corpus, but it is in no way an outlier.

52 Here, we use the nomenclature in textual criticism that distinguishes between works, version, and manifestation. See, e.g. D.C. Greetham, *Textual Scholarship: An Introduction*, 2nd ed. (New York: Routledge, 2015); Anna Gunder, *Hyperworks: On Digital Literature and Computer Games* (Uppsala: Uppsala University, 2004): 155–194.

53 Daniel Åberg, “At skrive til øret.”

54 Birgitte Stougaard Pedersen, “At skrive gennem lyd”, *Passage* 83 (2020): 90–91.

55 Anna Öqvist Ragnar, Storytel, zoom interview with first author, January 28, 2021.

56 See, for instance, John Feather, *A History of British Publishing* (London: Croom Helm, 1988); David Finkelstein and Alistair McCleery, *An Introduction to Book History* (New York: Routledge, 2005); Simon Eliot and Jonathan Rose, “Introduction”, in *A Companion to the History of the Book*, ed. Simon Eliot and Jonathan Rose (Oxford: Wiley-Blackwell, 2007).