Samlaren

Tidskrift för forskning om
svensk och annan nordisk litteratur
Årgång 140 2019

Apples and Oranges?

Large-Scale Thematic Comparisons of Contemporary Swedish Popular and Literary Fiction

BY KARL BERGLUND, MATS DAHLLÖF
& JERRY MÄÄTTÄ

Introduction

It is an age-old truth that you shouldn't compare popular fiction with literary fiction. The two categories are simply two very different kinds of literature.[1] Highbrow and lowbrow, if you will. Or literature with significantly different assets of cultural capital, to use the widespread notion of Pierre Bourdieu.[2] An analysis of the textual differences between popular and literary works of fiction risks just proving truisms or becoming irrelevant. You can't compare apples and oranges, right?

But what if one were to compare them anyway? What would one find? The previous objections might be true for more established comparative methods of literary studies, where you need to invest significant amounts of reading time and approach the titles one by one. If we compare some principal textual features of these two 'kinds' of literature on a larger scale, however, other, more unexpected, patterns might emerge. Are the popular and the literary such opposites, as preconceived opinions within the literary community tend to hold? What differentiates each category, and what do they have in common?

The aim of this article is to compare thematic trends in contemporary Swedish bestselling and literary fiction with the help of a computational method that extracts content themes based on statistical patterns of word usage in a digital text collection.[3] Furthermore, the aim is to discuss how the differences and similarities we find can be understood in relation to the socio-cultural contexts that surround these works of fiction, above all regarding literary prestige, author gender, and book-trade mechanisms. As far as we know, there are no previous studies that similarly compare the content of bestsellers and prestigious fiction on a larger scale.[4]

There are many ways in which computers allow us to compare different categories of text. We will focus on thematic aspects by applying *topic modelling* to our material. We consider themes, in the content-related sense, to be more interesting than, for in-

stance, style or narrative aspects because they reflect in a deeper way what literary texts are about. Do bestsellers recount police work to a higher extent than literary fiction? Probably. But how about family life? Or outdoor settings? Or strong emotions? One could guess, of course, but even with extensive knowledge of both popular and literary contemporary fiction, the answers would not be obvious. This article will provide answers to such questions by coupling quantitative textual analysis with an interpretative step based on qualitative domain expertise in contemporary Swedish popular and literary fiction. The overall purpose, as stated above, can be broken down into the following more concrete research questions:

- What are the thematic differences and similarities between contemporary bestsellers and literary fiction? How can these patterns be understood?
- What can the similarities between the popular and the literary tell us about contemporary fiction in general?
- Does literary fiction have recurring or typical themes in the same ways as genre fiction? Can literary fiction be understood as a genre in a thematic sense?
- Do parameters such as genre (e.g. crime fiction) or author gender have a stronger impact on thematic content than chart positions or literary prestige? How can the relationship between textual and extra-textual factors be interpreted regarding thematic content on a larger scale?

Our intention is to identify thematic differences and similarities between the popular and the literary which are systematically supported by empirical evidence. Hopefully, the article will also provide insights into what (thematically) characterises bestsellers and literary fiction.

Theoretical background

The article draws upon two theoretical domains: distant reading and computational criticism, on the one hand, and sociology of literature and book history, on the other. Franco Moretti's term *distant reading* not only challenged presupposed methodological practices within literary scholarship, it also highlighted how distance, as opposed to close reading, itself changes literary analysis:

> Distant reading: where distance, let me repeat it, is *a condition of knowledge*: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems.[5]

An analysis based on topic modelling is a good example of this. It gives us a condensed view of a whole, possibly very large, collection of texts, but at the same time assigns topics to individual words. It represents a mode of analysis which is not within the reach of human close reading, but whose output can reveal new patterns which are comprehensible to the literary scholar and which invite an interpretation.

Scholarship in the tradition of Moretti has mostly focused on textual analysis and paid little attention to extra-textual factors. Such work has occasionally even been labelled as "quantitative formalism".[6] This disregard for context is problematic for sociologists of literature and book historians, who emphasise the importance of authors, marketing, reception, literary institutions, etc.[7]

Only recently have researchers engaged in large-scale computational criticism started to address this lack of contextual awareness. Andrew Goldstone points out that the sociological perspectives that were a part of Moretti's original concept have faded away in later scholarly practice. Instead, and according to Goldstone due to the strong tradition of text interpretation within literary scholarship, the "reading" part in distant reading has been over-emphasised and thus transformed the concept into yet another version of such an interpretative practice, albeit with more texts taken into account.[8]

Ted Underwood credits the long pre-history of "distant readings" before Moretti and emphasises the problems with seeing distant reading as a plain digital humanities application. For Underwood, it is a perspective on literary history that is productive not because of its use of computers, but due to its systematic approach, its links to social sciences methods, and its integration of an experimental setting:

> My rationale is simple. An approach to literature informed by social science can produce significant historical results by itself—with or without computers. But the converse has not generally turned out to be true. Computational methods, by themselves and without a social scale of inquiry, have not been enough to transform literary history.[9]

Katherine Bode, finally, makes a similar claim when she criticises the Moretti tradition of being "inherited from, not in opposition to, the New Criticism and its core method of close reading".[10] Instead, she seeks to interconnect what she labels as *data-rich literary history* to the activities and insights from textual scholarship, "the bibliographical and editorial practices that literary scholars have long relied on to interpret and represent the historical record".[11]

The perspectives in this article resemble those of Goldstone, Underwood, and Bode, but are also highly indebted to the long and vivid tradition of Swedish sociology of literature, which goes back at least to the 1960s. While defined and discussed many times over the years, Johan Svedjedal's definition captures the key idea of this perspective: "to systematically analyse the relationships between fiction and society."[12] Studies in this

tradition have ranged from investigating the ways in which fiction depicts and affects society to focussing on, for instance, the book market, literary criticism, and readers. Some of the main characteristics of these studies have been a materialist perspective, group perspectives, a broad view of what constitutes literature, and an unwillingness to pass value judgements.[13]

We thus seek to theoretically unite large-scale computational perspectives on textual corpora with contextually well-founded sociology of literature.[14] Instead of taking a really large dataset as our point of departure—a common practice in distant reading, where 'the bigger, the better' has for long been a standard assumption—we take as our point of departure a conceptually well-defined and carefully curated corpus.

The composition of our corpus is key to what we can reveal about contemporary literary culture, economy, and prestige (see more on the corpus below). By setting up an experiment that enables thematic comparisons between the popular and the literary on a broader scale, we can hopefully provide an empirically well-founded theoretical discussion on the relationship between textual content and extra-textual factors in relation to contemporary popular and literary fiction.

A computational approach is necessarily to some extent systematic and will not be biased by implicit factors. Experiments are based on explicit assumptions and data, and the validity of conclusions can be tested by attempts to replicate the results. Such approaches have been criticised, however, for their reductive perspective on literature. Sarah Allison's notion of *reductive reading* is a response to this objection. Allison points out that all readings—even close ones—are reductive in one way or another, but only readings that use computational methods explicitly lay bare what is taken into account in the literary analysis. A computational approach prompts the literary scholar to think through and motivate the methodical choices made when defining the analytical concepts in quantitative terms.[15] Thus, a reductive approach in large-scale literary modelling can provide a solid foundation for further qualitative (and more nuanced) discussions. Or, as Allison frames it: "Reductive reading, by contrast, clears space for reading that is *not* reductive."[16]

So, how does one set up such an experiment?

Experimental setup: Material and method

Our approach involves three components. First, there is the curation of a corpus of literary texts. Secondly, we extract a number of topics by means of a statistical analysis of the corpus. Thirdly, we make an interpretation of these topics, based on our domain expertise in contemporary Swedish fiction. This section describes parts one and two, while the rest of this article is dedicated to the third, interpretative, step.

CORPUS

A study on popular and literary fiction needs to operationalise these categories in a suitable way. We take the Swedish bestseller charts and the titles shortlisted for the August Prize (Augustpriset), one of the major Swedish literary awards, as our point of departure.

The *popular corpus* comprises all original Swedish prose fiction on the annual bestseller charts compiled by the book-trade journal *Svensk Bokhandel*, from the start of the annual lists in 2004 up to 2017, inclusive, and in the two most commercially important categories: hardbound and paperback. These lists are based on point-of-sale data from booksellers. They do not represent all book sales in Sweden but cover the major share. Earlier studies have considered the lists to be reliable for tracking bestselling fiction in Sweden.[17]

With duplicates removed, 182 of these 186 novels (98%) were at the time of writing available in digital format and thus form the corpus. Of these bestsellers, as many as 125 (69%) are works of crime fiction.[18] There are, of course, other genres prevalent in the corpus, such as chick lit/romance, feelgood, and autofiction, but they are all more uncommon and hence not systematically separated in our experiments.[19] The gender distribution is more or less even (46% works by female authors, 49% works by male authors, 5% works by mixed-gender authors).

The *literary corpus* is based on titles shortlisted for the August Prize. The prize was created in 1989 by the Swedish Publishers' Association (Svenska Förläggareföreningen) with the explicit aim to market Swedish literary fiction. Since the late 1990s it has arguably been the most important and prestigious literary award for Swedish literature. It was modelled on the Booker Prize (which was in turn modelled on the Goncourt Prize) and follows a similar procedure, in that a jury first selects a shortlist comprising six titles. Where it differs from most other literary prizes and awards is that the winner is then selected by a secret group of so-called *electors* consisting of an equal number of booksellers, librarians, and literary critics (among others).[20] Since 1992, the August Prize has been awarded in three categories—adult fiction, adult non-fiction, and children's and YA fiction and non-fiction—each with its own jury, shortlist, and group of electors.[21]

The corpus consists of all prose-fiction titles shortlisted for the Swedish August Prize in its category for adult fiction between 2004 and 2017, inclusive, that were digitally available at the time of writing. The 58 titles in the corpus comprise more than two thirds (69%) of all the 84 titles shortlisted in the category for adult fiction between 2004 and 2017, almost all of the prose-fiction titles (58 out of 68, or 85%), and they include all but one of the winning titles.[22] This corpus has a more or less even gender distribution—26 of the titles (45%) are written by female authors.

Even though the August Prize is of clear commercial relevance, and the inclusions and omissions of particular shortlists are frequently debated (increasing public interest in the prize), the books nominated for the prize through the years have often been by some of the most famous and prestigious authors of the Swedish literary field of the last few decades.[23] The titles range from historical novels to contemporary realist novels and autofiction. Only a few of them could be said to border on genre fiction.[24] Interestingly, very few (if any) shortlisted titles show affinities with crime fiction, while, conversely, a substantial number of the shortlisted titles build on earlier works of canonical literature, minor classics, or the lives of important writers and thinkers.

One of the main functions of major literary prizes and awards such as the Booker Prize, the Goncourt Prize, and the Swedish August Prize is to generate media interest, benefiting not only the prestige of the authors, but also increasing their sales.[25] In fact, important prizes and awards often function as creators of what French sociologist of literature Robert Escarpit called *blockade runners* (*forceurs de blocus*), that is, books that break down the social barriers or even blockades between popular fiction and literary fiction. In this case, the prize makes certain literary fiction titles visible, accessible, and (socially) attractive to a large reading and book-buying audience.[26] Another way of looking at it is through Hans Hertel's concept of the *media elevator* (*medieelevator*), where books and stories first published for a relatively small audience can, through marketing, media attention, adaptations, etc., reach an ever-larger distribution, all the way to becoming orally distributed, even leaving the book medium and any adaptations behind.[27]

As the August Prize has been very successful in increasing the sales of especially the winning titles, more than two thirds of the winning titles in the corpus (9 of 13, or 69%) are found on the bestsellers charts in *Svensk Bokhandel*, as are a handful (5 of 45, or 11%) of the other shortlisted titles in the corpus. This means that there is an overlap between the two corpora, consisting of 14 titles, or almost a tenth (8%) of the 182 titles in the popular corpus, and almost a quarter (24%) of the 58 titles in the literary corpus.[28] Our operationalisation of popular fiction and literary fiction is thus not as two mutually exclusive categories. This reflects how the popular and literary domains often interconnect in the contemporary book trade.[29]

To summarise, we use titles on Swedish bestseller charts as a proxy for popular fiction, and titles shortlisted for the August Prize as a proxy for prestigious literary fiction. Since crime fiction is so predominant in the charts, we will also single out the genre and thus compare non-crime fiction bestsellers to titles shortlisted for the August Prize. The structure of our material is as follows:

Titles on bestseller charts: 182
   (125 of which are crime fiction and 57 non-crime fiction)
Titles shortlisted for the August Prize: 58
Overlaps (shortlisted bestsellers):   14

DISTANT READING BY MEANS OF TOPIC MODELLING

Our variety of "distant reading" investigates what content themes are present in literary texts and how they distribute over different categories of text. The formal tool we make use of is topic modelling. This is a statistical method whose purpose is to uncover latent themes (topics) that can be found in a corpus of text pieces. In our set-up, we have decided to regard paragraphs as the basic units. Paragraphs can be assumed to be thematically coherent and they are a 'natural' category in prose fiction.[30]

The topics are components of an abstract mathematical model of how paragraphs are generated by a process of random decisions. A certain number of topics are involved: This number is selected in the experiment set-up. Each paragraph can (of course) be viewed as a sequence of words. The idea is that the words are generated in such a way that we first draw a topic according to a statistical distribution of topics associated with the paragraph, and then randomly select each word given a distribution of words associated with that topic.[31] However, as is common practice, we ignore some words (very common and very rare ones).

The hypothesis behind the application of topic modelling is that the distribution of topics in a paragraph corresponds to its mixture of content-related themes (for instance, as we will see, *meals* and *outdoor scenery*), and that words are more or less strongly associated with such themes. The modelling process consists in determining a model (formed by the two statistical distributions) that to a high extent agrees with the actual paragraphs of the corpus.

The current approach to topic modelling is (apart from the decision to use paragraphs as the units to be modelled) similar to a setup we have worked with before, in a comparison of Swedish classical literature and more recent bestsellers.[32] So, our set-ups are justified by previously having been able to generate meaningful and interesting analyses for similar data.

The words have been assigned a part-of-speech tag.[33] Lexical terms are formed by combining the base form and the part-of-speech tag. Inflected forms are in this way grouped as one term (lemma). The part-of-speech tag will disambiguate some lemmas. Only noun and verb lemmas are fed to the topic modelling, and the 100 most frequent lemmas are removed.[34] There is also a requirement that each term must be found in at least 20 books. This condition eliminates words which are specific to a few books. Actually, most words are unusual in this sense.

We use a version of Latent Dirichlet Allocation as implemented by the Mallet software.[35] It is guided by the same parameter settings as in our previous study.[36] We should stress that we use topic modelling in an explorative and unsupervised fashion. The previous study has shown, and the present one confirms, that our procedure generates topics that are intuitive to interpret and make sense to scholars of literature.

The generated model assigns a topic (identified by a number) to each word token (included in the data). We can consequently express how common a topic is in a work or a subset of works as its relative frequency, and then relate subsets to each other by calculating the ratio[37] between their relative frequencies (e.g. the relative frequency of the bestseller corpus divided by the relative frequency of the literary corpus).

The plain relative frequency of a topic gives larger works more weight. Large texts may consequently "distort" relative frequencies from the point of view of their use to characterise a subset. To avoid that effect, we also take a rank-based look at the data. This makes it possible to record the probability that a randomly selected work of a given category contains more of a certain topic than a randomly selected work outside of that category (but from the main corpus). We call this the *outranking ratio* of a subset of the corpus with respect to a topic. A value well above 50% indicates that the topic is characteristic for the category, and correspondingly for lower numbers. Numbers around 50% show that the category is "neutral" with respect to the topic in relation to the corpus as a whole. Thus, the relative frequency ratios have the advantage of allowing straight comparisons between the collected text of subsets, while the outranking ratios give equal weight to short and long works. We consider both relative frequency ratios and outranking ratios in our analysis.

As is common practice, we present the topics for "reading" as lists of "top" keyword terms. The keyword ranking is based on the chi square value, which quantifies the strength of association between a term and a topic.[38] The raw topic modelling results, on which we base our analysis, are available online as supplementary material.[39]

It should be emphasised that our conclusions are not to be seen as objective hard data points. Our analysis is highly qualitative, since it depends on our interpretation of the results. Hence, the method is in many ways surprisingly similar to how literary scholars in general tend to work, but with the main point of departure as a crucial difference: Instead of analysing works of literature, we analyse statistical output from an experiment where we have modelled works of literature. Thereafter, we relate the output to literature, trying to be conscious of and upfront about the limitations of the method. Nevertheless, the method lets us see new things, or old things in new ways, both on the level of subsets, genres, and other categories and on the level of individual works.

Comparing bestsellers and shortlisted literary fiction

If the titles on the Swedish bestseller charts are compared straight away with the short-listed literary titles, a relatively clear pattern emerges, on all levels of thematic specificity. When k=20, i.e. when the model sorts all words into 20 topics, thus producing broad themes, the clearly more common topics in the bestsellers cover themes such as "police work and murder investigations" (topic 10), "telephone calls" (19), and "money and business" (20), but also more action-related phenomena such as "interiors/moving inside buildings" (7), "demands and challenges" (2), and "driving, cars, and traffic" (13):

> k=20, topic 10; ratio Bs/Aug: 2.70; outranking ratio: 83%
> polis, mord, utredning, förhör, offer, mördare, kollega, åklagare, gärningsman, brott
> [*police, murder, investigation, interrogation, victim, murderer, colleague, prosecutor, perpetrator, crime*][40]

> k=20, topic 19; ratio Bs/Aug: 1.54; outranking ratio: 81%
> ringa, telefon, svara, samtal, mobil, nummer, be, läkare, lur, mobiltelefon
> [*dial, telephone, answer, call, mobile, number, ask, doctor, handset, mobile phone*]

> k=20, topic 20; ratio Bs/Aug: 1.36; outranking ratio: 68%
> peng, betala, sälja, köpa, krona, miljon, affär, företag, kosta, bank
> [*money, pay, sell, buy, Swedish krona, million, business/shop, company, cost, bank*]

> k=20, topic 2; ratio Bs/Aug: 1.23; outranking ratio: 77%
> behöva, böra, försöka, hjälpa, skull, ge, tid, hoppas, hitta, klara
> [*need, should, try, help, sake, give, time, hope, find, manage/cope with*]

> k=20, topic 13; ratio Bs/Aug: 1.25; outranking ratio: 62%
> bil, köra, väg, gata, svänga, stanna, passera, buss, tåg, parkera
> [*car, drive, road, street, turn, stop, pass, bus, train, to park*]

> k=20, topic 7; ratio Bs/Aug: 1.11; outranking ratio: 65%
> dörr, rum, öppna, vägg, fönster, trappa, stänga, hall, korridor, golv
> [*door, room, open, wall, window, staircase, close, hallway, corridor, floor*]

With a larger number of topics, the picture remains more or less the same, apart from several over-represented topics of a more specific nature turning up.[41]

The topics more common in literary fiction have a bias toward themes concerning "nature, water, and weather" (topic 6), and "existential reflection and strong emotions" (1), but also, to some degree, topics addressing "family circles" (9), "state affairs, war, history, and politics" (8), "clothes and appearance" (15), and "body parts, corporeality, intimacy, and violence" (3):

k=20, topic 6; ratio Aug/Bs: 1.61; outranking ratio: 76%
sol, vatten, hav, himmel, vind, träd, båt, strand, ljus, skog
[*sun, water, sea, sky, wind, trees, boat, beach/shore, light, forest*]

k=20, topic 1; ratio Aug/Bs: 1.46; outranking ratio: 73%
liv, människa, känsla, leva, kärlek, död, värld, minne, älska, uppleva
[*life, human, emotion, to live, love, death, world, memory, to love, experience*]

k=20, topic 8; ratio Aug/Bs: 1.25; outranking ratio: 64%
land, krig, president, tal, kung, svensk, tysk, regering, fiende, prins
[*country/nation, war, president, speech, king, Swedish, German, government, enemy, prince*]

k=20, topic 9; ratio Aug/Bs: 1.19; outranking ratio: 54%
mamma, barn, pappa, familj, mor, far, son, förälder, bo, flicka
[*mum, child, dad, family, mother, father, son, parent, to live/inhabit, girl*]

k=20, topic 15; ratio Aug/Bs: 1.18; outranking ratio: 60%
hår, kläder, skjorta, sko, klänning, jacka, tröja, byxa, bära, hänga
[*hair, clothes, shirt, shoe, dress, jacket, sweater, trousers, wear, hang*]

k=20, topic 3; ratio Aug/Bs: 1.14; outranking ratio: 57%
kropp, arm, blod, finger, bröst, kind, mun, mage, andas, ansikte
[*body, arm, blood, finger, breast, cheek, mouth, stomach, breathe, face*]

The tendencies appear to be similar when a larger number of topics are extracted, although more specific themes gain in prominence.[42]

Some expected thematic differences between the popular and the literary emerge here. It is not surprising, for instance, that themes related to police work, murder investigations, shootings, and criminal procedure are more prevalent among the bestsellers. These can all be connected to the genre of crime fiction, which makes up roughly two thirds of the bestseller corpus.

The topics that are prominent for literary fiction are hardly unexpected, but they are not as easy to predict as the themes related to crime fiction. One could possibly have guessed that themes of existential reflection, intellectual development, and speech and language would be more common in literary fiction than in popular fiction. Nevertheless, it is interesting to see this kind of thematic characterisation of the literary realm, since it is a type of fiction that literary critics usually refuse (or find it hard) to nail down thematically. Ever since the Romantic era, the literary ideal has been uniqueness and originality, not imitating other writers and certainly not making use of genre conventions or other kinds of textual clichés. However, our analysis indicates that there seem to be themes typical of contemporary literary fiction, or at the very least Swedish contemporary fiction shortlisted for the August Prize. (The themes shunned by literary fiction are, of course, also an important part of its identity.)

COMING OF AGE

Many topics over-represented in the literary corpus appear to deal with childhood or adolescence in different ways: generically (e.g. k=20, topic 9; k=40, topic 34; k=80, topics 56, 69), concerning the fears and struggles of how one should live one's life (e.g. k=20, topic 1; k=40, topics 9, 13; k=80, topic 1), or how to break away from something, perhaps through education (k=40, topic 4), love (k=40, topic 11; k=80, topic 9), or conquering language (k=80, topic 67). Thus, the coming-of-age story, more or less. Overall, the literary realm seems to be more retrospective than the popular; it deals with family trouble rather than work trouble and is more focused on psychological conflicts and strong emotions connected to love, death, and remembrance of things past.

Prominent examples of often difficult family and childhood depictions in the literary corpus include Tomas Bannerhed's *Korparna* (2011), Göran Rosenberg's *Ett kort uppehåll på vägen från Auschwitz* (2012), Per Olov Enquist's *Ett annat liv* (2008), and *Liknelseboken* (2013), and Carola Hansson's *Masja* (2015), as well as corpus-overlapping titles such as Susanna Alakoski's *Svinalängorna* (2006), Jonas Hassen Khemiri's *Montecore* (2006), and Åsa Linderborg's *Mig äger ingen* (2007). Many of these also deal with themes such as language and writing, existential reflection, and the difficulties of love and living one's life.[43] Again, these are not astonishing results, but clear and interesting ones.

PLOT VS. SETTING

Furthermore, it makes sense that the bestsellers show an over-representation in topics such as telephoning, driving, moving inside buildings, transportation, and dialogue (see below). Generally speaking, popular literature is, to a greater extent than literary fiction, occupied with the plotline; things that—in a very concrete way—happen is what generates the reading pace. At the other end of the line, literary fiction seems to indulge in themes mostly concerning setting: nature, weather, water, skies, colours, light, lighting, etc. For instance, when k=80, three of the seven most over-represented topics (46, 45, 65) all show content of this kind.

In fact, this contrast between topics related to plotline and events and topics related to settings is one of the most apparent in the whole comparison. Hence, popular fiction focuses upon what happens, literary fiction upon depiction and setting; popular fiction might even be characterised as urban, up-tempo, and contemporary, literary fiction as pastoral, serene, and focused on relatively recent history. This is hardly news or unexpected, but the quantitative strength of this pattern is, at least to us, a bit of a surprise.

LITERARY SURPRISES

More surprising is that the topic "state affairs, war, history, and politics" is over-represented in the prestigious literature. At first glance the topic—with top words such as "president", "ubåt" [*submarine*], "soldat" [*soldier*], "ryss" [*Russian*] etc.—seems to indicate exactly the kind of thematic content common in high-paced thrillers with an international setting—in Sweden for instance by bestseller writer Jan Guillou. This assumption proves to be at least partly right, since Guillou is indeed the author with the highest representation in this topic.[44] But apart from Guillou (and some other bestsellers by writers such as Jonas Jonasson and Leif GW Persson), the topic is generally more prevalent in the literary realm; the novels with the least representation of the topic are to a high extent bestsellers. At least in a Swedish context, then, and if our corpuses are fair operationalisations, state affairs, war, history, and politics are thematically atypical to popular fiction.[45] In general, the topic seems to relate more closely to literary fiction, with high representation in the works of authors such as Malte Persson, Göran Rosenberg, Peter Fröberg Idling, Jörgen Gassilewski, and Per Olov Enquist.[46]

An equally surprising find is that body parts, corporeality, intimacy, and violence are to a greater extent connected to the literary domain than to bestsellers. The general assumption would probably be that sex and violence are essential parts of popular fiction, especially crime fiction and romance. Our investigation indicates that this prejudice might not hold true, or at least that it needs to be nuanced regarding crime fiction.[47]

If one examines the theme a bit closer, the top words highlight all kinds of connections to the body, both good and bad, both intimate and violent.[48] Hence, the topic absolutely covers sex, but also violence. One would think that such a theme would show a strong affinity with crime fiction—but it would not seem so. If we look at the top titles in this topic, an interesting pattern emerges. We find literary fiction (titles by e.g. Carolina Fredriksson, Tomas Bannerhed, Amanda Svensson, Carl-Michael Edenborg, and Lyra Ekström Lindbäck) interspersed with some bestsellers, most notably several titles by Lars Kepler. The bottom segment, on the other hand, consists mostly of bestsellers, primarily crime fiction, historical novels by Jan Guillou, and feelgood novels by Jonas Jonasson.

Kepler piques curiosity in this context, since it is a pseudonym for Alexander Ahndoril and Alexandra Coelho Ahndoril, who both, before they started to write crime fiction together, wrote literary and award-winning fiction. As writers, they have thus, sociologically speaking, moved from the literary realm to the popular. Furthermore, Kepler is commonly known for crime fiction exhibiting a lot of violence. Because of this, they have occasionally been accused by critics of pushing the genre in an unwanted direction.[49] Our findings seem to indicate that Kepler are an outlier in the crime fic-

tion genre in their use of graphic violence. Hence, Kepler might not have pushed the genre of crime fiction towards its own extremes, but rather—in a crime fiction narrative—incorporated a more literary vocabulary for the portrayal of bodies, sex, and explicit violence.

Contemporary Swedish crime fiction certainly contains violence—the genre's foundation is, of course, murder and investigation—but perhaps not so much graphic depiction of either violence or sex as has been previously assumed.[50] A possible explanation for the result could be that literary fiction tends to linger on subjective corporeal experiences of sex and violence, whereas crime fiction might often prefer to avoid awkward descriptions of sex and to gloss over the subjective pain of its protagonists and victims, and so leans more on objective descriptions. Either way, this is something that might at least partly challenge common perceptions of crime fiction (and literary fiction).

INDIVIDUAL EXCEPTIONS

Most of these observations concern general patterns that thematically differentiate the popular and the literary, but there are, of course, titles and authors that break these patterns. Jan Guillou's affinity with the literary-marked "state affairs" theme has already been mentioned, but there are others. Jörgen Gassilewski's *Göteborgshändelserna*, shortlisted for the August Prize in 2006, has a high representation of the bestseller-marked topic of police work and murder investigations (k=20, topic 10); several novels by Johan Theorin are in the top segments in the clearly prestige-marked topics of outdoor settings (k=20, topic 6; k=40, topic 21; k=80, topic 45); both *Skam* and *Skugga* by bestseller author Karin Alvtegen are among the highest ranking on the literary-marked topic of existential reflection and mental states (k=20, topic, 1); the 2015 August Prize winner *Allt jag inte minns* by Jonas Hassen Khemiri has a high representation of the bestseller-marked topic of telephoning (k=20, topic 19); August Prize nominee *Välkommen till Amerika* by Linda Boström Knausgård has a high representation of the bestseller-marked topic of entering and moving inside buildings (k=20, topic 7); et cetera.

All such thematic blockade runners have plausible explanations: Gassilewski's *Göteborgshändelserna* portrays the demonstrations, riots, and police brutality during the 2001 EU summit in Gothenburg; the Swedish island Öland and its idyllic nature, popular among both Swedish and foreign tourists, plays an important role in all crime novels by Theorin, as well as in the marketing of them;[51] Karin Alvtegen writes psychological thrillers that border on the literary realm;[52] telephone calls between the protagonist and a subsidiary character play an important part in Khemiri's *Allt jag inte minns*; to a great extent, Boström Knausgård's *Välkommen till Amerika* takes place inside an apartment, where the main characters move back and forth between the rooms.

Interestingly, however, while the exceptions from the literary domain seem to be just individual exceptions, where a theme of a certain novel happens to coincide with themes prevalent among the bestsellers, the exceptions from the popular domain are different. These latter are by writers that have, in one way or another, tried to approach a more prestigious writing—this holds true for Alvtegen and Theorin as well as for Guillou. These exceptions do not concern single titles but several, thus indicating that these authors in fact lie closer to the literary realm when it comes to certain thematic aspects (existential reflection for Alvtegen; outdoor settings for Theorin; state affairs, war, history, and politics for Guillou). The same cannot be said about Gassilewski, Khemiri, or Boström Knausgård, who are all high-prestige authors far from the popular realm (although Khemiri is more broadly popular and shows up on bestseller charts).

THE GENDER DIMENSIONS

Moreover, there seems to exist a negative correlation between some topics. For instance, works with high levels of state affairs, war, history, and politics have particularly low levels of body parts, corporeality, intimacy, and violence. This might be a case of opposite thematic poles that transcend literary status and genre: Novels either contain an abundance of historical facts, war, and politics, or intimacy, sex, and (bodily, graphic) violence, but rarely both. Is it the narrative difficulties of combining macro history with micro history that shows up here?

What these poles do connect to is the gender dimension: men write more on the theme of "state affairs and history" (ratio: 1.84, the most male-biased topic when k=20), women more on the theme of "body and intimacy" (ratio: 1.25, the fourth most female-biased topic when k=20; the two most female-biased topics concern "eating and drinking" and "appearance and clothing"). There are presumably also other parameters at work, waiting to be discovered.

The rule-breakers can also be identified on a gender axis. In the male-biased topic of state affairs, war, history, and politics, Kristina Ohlsson's crime novel *Paradisoffer* (2012) and Therese Bohman's campus novel *Aftonland* (2016) have very high representation, while Tom Malmquist's autobiography on the loss of his wife, *I varje ögonblick är vi fortfarande vid liv* (2015), and John Ajvide Lindqvist's vampire novel *Låt den rätte komma in* (2004) are in the bottom extremes. In the female-biased topic of bodies, corporeality, intimacy, and violence, *Låt den rätte komma in* and Mikael Niemi's crime and love novel *Mannen som dog som en lax* are amongst the highest ranked, while Lena Andersson's love novel *Egenmäktigt förfarande* and Ohlsson's *Paradisoffer* are in the bottom segment.

This outcome make sense when looking at the titles one by one. Still, the results are significant: Kristina Ohlsson, Therese Bohman, and Lena Andersson can all be said

to write in a more stereotypically "male" way (informative, dense, low on emotions), while Tom Malmquist, John Ajvide Lindqvist, and Mikael Niemi, by contrast, deal heavily with bodily experience.

In this way, a topic modelling approach enables investigations of thematic norms in subsets, and of outliers from these norms, both concerning literary prestige and author gender. Instead of reasoning merely about discrete categories—such as bestsellers or shortlisted literary fiction, works by male or female authors—we can make use of these (and similar) binaries to produce continua, ranges, spans. In line with Ted Underwood, we argue that this is a factor that makes statistical modelling especially suitable for the analysis of literature, with its by-nature blurry lines.[53] In our model, a novel found on the bestseller charts can have literary qualities (statistically speaking); a title by a male author can show congruence with what distinguishes female writing (statistically speaking). If we proceed from calculation, then themes, genres, authors, and novels can all be arranged and analysed based on their respective affinities. Despite what many scholars in the humanities believe, numbers are useful for displaying ambiguity.[54]

Comparing non-crime fiction bestsellers and shortlisted literary fiction

Since crime fiction is so predominant among the bestsellers (just over two thirds), the genre of course has an enormous impact on all the comparisons above. To a great extent, what we find are differences between commercially successful crime fiction and literary fiction. If we exclude all crime fiction from the bestseller corpus and instead compare non-crime fiction bestsellers with shortlisted literary fiction, we find both similarities and differences.

First, the thematic differences between the sub-corpora are generally smaller. Where the extreme topic "police work and murder investigations" (k=20, topic 10) has a ratio of 2.70 when bestsellers are compared straight away to the shortlisted titles, the extreme for the same comparison with non-crime fiction, "money and economy", only has a ratio of 1.41 (k=20, topic 20). This outcome is expected, since crime fiction is far from the prestigious realm.[55]

Secondly, several broader topics over-represented among the bestsellers generally are also over-represented among the non-crime fiction bestsellers, although to a lesser extent. This holds true for "money and economy", "demands and challenges", "telephone calls", and "police work and murder investigations" (k=20, topics 20, 2, 19, and 10), although the latter drops from the first to the fifth position and now only has a slight bias towards the popular. When the number of topics is larger, the same goes for some similar, but more fine-grained, themes as those mentioned above.[56]

FOOD AND MONEY VS. NATURE

Regardless of the number of topics chosen in the model, there are always over-represented topics connected to "money and economy" and "telephoning".[57] These themes thus appear to be equally biased for bestsellers and non-crime fiction bestsellers when compared to prestigious literature. In fact, topics relating to "money and economy" are even more biased in the non-crime fiction sub-corpus.

Another apparent trend is the general over-representation of topics connected to eating and drinking:

> k=20, topic 18; ratio BsNonCr/Aug: 1.31; outranking ratio: 74%
> dricka, kaffe, glas, äta, mat, vin, flaska, kopp, bord, hälla
> [*to drink, coffee, glass, eat, food, wine, bottle, cup, table, pour*]

> k=40, topic 39; ratio BsNonCr/Aug: 1.42; outranking ratio: 70%
> glas, dricka, vin, flaska, öl, gäst, klunk, bar, beställa, restaurang
> [*glass, to drink, wine, bottle, beer, guest, sip, bar, order, restaurant*]

> k=80, topic 66; ratio BsNonCr/Aug: 1.40; outranking ratio: 73%
> bord, kaffe, bjuda, kopp, gäst, restaurang, tant, matsal, middag, te
> [*table, coffee, to offer, cup, guest, restaurant, aunt, dining room, dinner, tea*]

> k=80, topic 75; ratio BsNonCr/Aug: 1.25; outranking ratio: 66%
> dricka, glas, vin, flaska, öl, hälla, klunk, bar, alkohol, champagne
> [*to drink, glass, wine, bottle, beer, pour, sip, bar, alcohol, champagne*]

> k=80, topic 57; ratio BsNonCr/Aug: 1.21; outranking ratio: 74%
> äta, mat, laga, bröd, tallrik, n[58], mjölk, kök, tugga, middag
> [*eat, food, to cook, bread, plate, n, milk, kitchen, to chew, dinner*]

> k=160, topic 101; ratio BsNonCr/Aug: 1.33; outranking ratio: 78%
> äta, middag, lunch, frukost, matsal, smörgås, smaka, ost, bjuda, äpple
> [*eat, dinner, lunch, breakfast, dining room, sandwich, to taste, cheese, to offer, apple*]

> k=160, topic 43; ratio BsNonCr/Aug: 1.33; outranking ratio: 71%
> glas, vin, bord, kök, skål, kylskåp, diskbänk, hälla, champagne, gäst
> [*glass, wine, table, kitchen, bowl/toast, refrigerator, kitchen-sink, pour, champagne, guest*]

> k=160, topic 131; ratio BsNonCr/Aug: 1.29; outranking ratio 77%
> mat, laga, jul, bröd, fira, ägg, potatis, smör, korv, julafton
> [*food, to cook, Christmas, bread, celebrate, egg, potato, butter, sausage, Christmas Eve*]

These topics were not biased at all when crime fiction was included, and thus appear to be thematically important only in other kinds of bestsellers, such as novels by Fredrik Backman, Emma Hamberg, Helena Henschen, Kajsa Ingemarsson, Kristina Sandberg, and Alex Schulman.

On the other side of the literary spectrum, topics concerning setting are the most apparently biased ones, both on broader and narrower levels of specificity. On all levels, such topics are found in the top extreme: "outdoors (nature, water, weather)" (k=20, topic 6), "maritime activities and settings" (k=40, topic 21), "skies, weather" (k=80, topic 46), and "trees, gardens" (k=160, topic 52). In fact, when k=80, the three topics most biased towards literary fiction connect to settings, specifically weather, the outdoors, and nature:

> k=80, topic 46; ratio Aug/BsNonCr: 1.58; outranking ratio: 77%
> sol, himmel, regn, flyga, luft, moln, stjärna, frysa, värme, väder
> [*sun, sky, rain, to fly, air, cloud, star, freeze/be cold, heat, weather*]

> k=80, topic 45; ratio Aug/BsNonCr: 1.42; outranking ratio: 72%
> snö, sten, mark, träd, vind, jord, grav, gräva, gren, falla
> [*snow, stone, ground, tree, wind, earth, grave, dig, branch, to fall*]

> k=80, topic 61; ratio Aug/BsNonCr: 1.40; outranking ratio: 67%
> vatten, hav, båt, strand, ö, våg, brygga, hamn, simma, sand
> [*water, sea, boat, beach/shore, island, wave, pier, harbour, to swim, sand*]

There are, of course, other biased themes.[59] But taken together, the most manifest demarcation line between literary fiction and non-crime fiction bestsellers seems to be settings and depictions of the outdoors for literary fiction, and themes about money and food for popular fiction.

The prestigious realm excels in describing nature, even more so than when compared to bestsellers in general. A possible explanation is that crime novels set in picturesque rural areas by the sea—by writers such as Johan Theorin, Viveca Sten, Camilla Läckberg, and Mari Jungstedt—no longer weigh up on the popular side. At the other end, more prosaic themes about money, economy, and food appear to be appealing to the wider reading public.[60]

POETIC NATURE

If one breaks down this overall pattern and looks at the statistics for individual titles, top titles in topics concerning nature and outdoor settings include several novels by Sara Stridsberg, Carolina Fredriksson's *Flod* (2011), Tomas Bannerhed's *Korparna* (2011), Peter Törnqvist's *Kioskvridning 140 grader* (2009), Johannes Anyuru's *En storm kom från paradiset* (2012), Stina Stoor's *Bli som folk* (2015), and Cilla Naumann's *Springa med åror* (2012). These are all works that both thematically and conceptually portray nature and weather in different ways, be it because they are set in rural areas (Bannerhed, Törnqvist, Stoor, Naumann), have important scenes by rivers or in deserts (Stridsberg), or because they focus on themes such as homelessness (Fredriks-

son) or aviation (Anyuru). A common denominator for practically all of them is also that they deal with issues of social class, vulnerability, and alienation—which are, however, prevalent themes in Swedish literary fiction in general, but probably more difficult to find through topic modelling.

This pattern is not prominent, however, due to a handful of works being heavily focused on nature and the weather. Of the bottom 30 works in the two broader topics concerning nature (k=20, topic 6; k=40, topic 21), only two in total come from the prestigious realm: the very urban narratives in Jonas Hassen Khemiri's *Allt jag inte minns* and Lena Andersson's *Egenmäktigt förfarande*. Both these titles have, however, also appeared on the bestseller lists.

Hence, there seems to be a strong correlation between prestige and depictions of nature, the outdoors, weather, and rural settings, at least in Sweden. This might be linked to the prevalence of childhood narratives in Swedish literary fiction, often set in rural areas, and with a long tradition of often-idyllic Swedish nature poetry, but contrariwise also with a strong interest in issues of urbanisation and social class, where the countryside is associated with depopulation and poverty.

Possibly, one could instead argue that the reverse is true: Perhaps the broader Swedish reading audience is not all that interested in reading depictions of nature and weather, or the types of stories that tend to contain them. We cannot support such a claim with our numbers, but it would certainly be interesting to investigate the pattern further.

AND PROSAIC EVERYDAY LIFE

What a large audience of readers does seem to enjoy is reading stories containing scenes related to food and money. The former includes all kinds of topics, from broader ones to more specific themes such as "fika [having coffee] and restaurant visits" (k=80, topic 66), "cooking and eating at home" (k=80, topic 57), "bar scenes and drinking" (k=40, topic 39; k=80, topic 75), "domestic drinking" (k=160, topic 43), and what appears to be "eating together at feasts" (k=160, topic 131).

These themes are especially prevalent in bestsellers such as Kajsa Ingemarsson's *Små citroner gula* (2004), Emma Hamberg's *Mossvikenfruar: Chansen* (2005), Majgull Axelsson's *Jag heter inte Miriam* (2014), Fredrik Backman's *Britt-Marie var här* (2014), and Alex Schulman's *Glöm mig* (2016), and corpus-overlapping titles such as Kristina Sandberg's historical novels about the life of housewife Maj (2010–2014), Susanna Alakoski's *Svinalängorna* (2006), and Åsa Linderborg's *Mig äger ingen* (2007).

The highest-ranking titles—by Ingemarsson and Sandberg—are novels that all deal with food in some way: in Ingemarsson's chick-lit narrative, the protagonist is a restaurant worker; Sandberg's twentieth-century historical novels about housewife Maj are

to a great extent centred on her cooking and preparing meals for family and friends. In the other high-ranking titles, scenes that contain eating and drinking are simply commonly occurring and important for the narratives.

Interestingly, what seems to surface is chick lit and realist comedy, on the one hand (Ingemarsson, Hamberg, Backman), and sombre twentieth-century historical novels, on the other (Sandberg, Alakoski, Linderborg, Axelsson, Schulman). Food thus seems to be used to create either cosy or romantic settings, or be a part of prosaic, melancholy, or even harrowing ones. What goes for all of the above, however, is the strong connection to realism: Ingemarsson, Hamberg, and Backman all write light-hearted, entertaining stories about contemporary Sweden, well-grounded in detail concerning everyday life, whereas the novels by Alakoski and Linderborg are social-realist accounts of childhood and adolescence in working-class families with drinking problems; Schulman's novel deals with related themes, but in an upper-class milieu; Sandberg is proficient in the details of the work of housewives in 1930s to 1950s Sweden; and Axelsson interlaces historical milieus from the Second World War and Nazi concentration camps with depictions set in post-war Sweden.

Food, thus, connects to realism. Themes about eating and drinking seem to be deployed to set the scene and make it lifelike, convincing, and/or recognisable for the readers. And this literary method seems closely associated with the types of stories more appreciated by the wider reading audience rather than by literary critics and prize-awarding juries.

Money seems to work in a somewhat similar fashion. Among the non-crime bestsellers, top works in topics related to money and economy are Jonas Jonasson's *Mördar-Anders och hans vänner* (2015) and *Analfabeten som kunde räkna* (2013), Catharina Ingelman-Sundberg's *Kaffe med rån* (2012), Liza Marklund's *Asyl* (2004), Helena Thorfinn's *Innan floden tar oss* (2012), Jan Guillou's *Att inte vilja se* (2014), and Filip Hammar's and Fredrik Wikingsson's *Tårtgeneralen* (2009). This reflects a blend of contemporary comedy (Jonasson, Ingelman-Sundberg, Hammar & Wikingsson) and historical, realist, or even documentary novels (Guillou, Thorfinn, Marklund). Again: a mixture of lighter and heavier themes that all connect to realism.[61]

Some notes on category-neutral topics

To this point, we have put our focus solely on over-representations and thematic extremes. Naturally, however, not all topics are biased towards either popular or literary fiction. Some, or even several, appear to be more or less stable in size in all the sub-corpora investigated.

When k=20, four topics are close to evenly distributed, both when bestsellers in general and non-crime fiction bestsellers are compared with the shortlisted literary titles, and when works by male authors are compared with works by female ones (see Table 1). Thus, a fifth of the broad thematic categories identified by the topic modelling seem to be unaffected by factors such as popularity, prestige, genre, and author gender.

Upon closer examination of the topics in Table 1, two of them appear to be rather vague and tend towards narrative elements that are common to most forms of prose fiction.[62] Topic 14—which deals with sleeping, going to bed, and waking up, but also with time in general—is also logical in this context. Almost all protagonists sleep, although in some narratives this is more of a focus—especially when it comes to sleeping problems. And more or less all novels feature characters who make plans for the immediate future, in one way or another.[63]

Topic 17 is more unexpected. It concerns reading and writing, and is the one topic of the four that most closely resembles a classical literary "theme". One could have suspected that such a theme would be more apparent in literary fiction, but the results indicate no such bias. When we expand the number of topics to k=40 and k=80, more specific versions of both sleeping matters and reading and writing appear.[64]

Although there are certainly differences in how this topic is depicted, bestsellers thus seem to display both processes of reading and writing to the same extent as literary fiction.

POPULAR AND LITERARY WRITING ON WRITING

When examining the top works in topic 17, the bestsellers consist of several titles by crime writers whose protagonists (just as the writers themselves) are reporters, most notably Liza Marklund and Stieg Larsson. In these novels, the working life of an investigative journalist is constantly highlighted—which, naturally, revolves around both reading and writing. But there are also examples of bestsellers that deal more explicitly with reading and writing, i.e. where reading and/or writing is the very subject of the bestselling novel. Such examples include Kajsa Ingemarsson's *Bara vanligt vatten* (2009), in which the protagonist is a successful, but lonely and bitter, crime writer; Linda Olsson's *I skymningen sjunger koltrasten* (2014), where the depressed protagonist meets a graphic novelist and an antiquarian bookseller, whose company slowly makes her better; Kerstin Ekman's *Grand final i skojarbranschen* (2011), a seemingly autobiographical account of the author's experiences in the world of books and publishing; and Karin Alvtegen's already mentioned *Skugga*, in which the narrative revolves around a Nobel laureate in Literature, a stolen manuscript, and a murder.

The top literary works in topic 17 are of a somewhat similar nature. On the one hand, there are novels in which the protagonist is a person who is heavily engaged in

reading and writing, as the art historian in Therese Bohman's *Aftonland* (2016) or the literary scholar in Eva Adolfsson's *En liten historia* (2009). In these, reading and writing are not so much a theme chosen as a consequence of a chosen setting. On the other hand, we have novels that thematise writing, either as a fictive or an autobiographical account, for instance Magnus Florin's *Ränderna* (2011), where the writer protagonist (with the same name as the author) gets assigned by an artist to write a piece on some photographs; the renowned Swedish writer Per Olov Enquist's literary autobiography *Ett annat liv* (2008); or Jonas Hassen Khemiri's *Montecore* (2006), a mock biography about the childhood and adolescence of writer Jonas Hassen Khemiri, which deals with language and writing to a great extent.

Taken together, reading and writing appear to be thematically neutral in contemporary Swedish fiction, although there are certainly more nuanced differences in how this topic is depicted in bestsellers and literary fiction. Perhaps media-specific biases (books, newspapers, letters, computers, etc.) could be revealing as to what types of reading and writing are most prevalent in popular and literary fiction. Or perhaps the apples and oranges aren't always that different, after all.

Conclusions

The results in this study range from the highly expected to the rather surprising. It was obvious that we would find a high over-representation among the bestsellers of themes related to police work, murder investigations, shootings, and criminal procedure. That we did so attests to the accuracy of the method.

Moreover, the contrast between themes connected to plotline (telephoning, driving, moving inside buildings, demands and challenges, etc.) among the bestsellers and setting (nature, water, weather, light, colours, etc.) among the literary novels was anticipated, although this pattern emerged more strongly than we would have thought. That there exists a gap between plot-focused popular fiction and language-and-style-focused literary fiction—and that this could be seen as a problem because literary fiction is said to no longer attract a large audience—has been frequently debated in Sweden over the last few decades.[65] Our findings support the idea that such a gap exists: literary fiction indeed seems to engage less in plot when compared to popular fiction, at least when it comes to the novels shortlisted for the August Prize. And as the latter rarely belong to the most experimental part of the literary field, one could suspect that this gap would be even wider if more niche titles had been chosen for the comparison.

Then we have the results that were not really expected, but neither very surprising. Rather, when noted, they make sense and fit existing explanations relating to literary prestige and gender patterns.[66] These results include the bias of themes about nature,

existential reflection, and intellectual development in the literary corpus; about food and economy among the non-crime fiction bestsellers; about body, intimacy, and appearance among the female writers; and about state affairs, war, history, and politics among the male writers.[67]

This outcome is not astonishing, but still, we argue, useful. For one thing, such results provide quantitative support for earlier qualitative claims. At least in our contemporary Swedish material, male and female authors appear to give certain themes a systematically different amount of interest. Hence, these themes can be said to be biased to male and female concerns. Gender differences in writing is a subject that has always been a moot point. Hopefully, our results can inspire further investigations into the nature of these differences, and their causes. Here, we merely note that the most gender-biased themes align quite clearly with traditional gender norms.[68]

Secondly, our results can provide brief thematic outlines also for genres and subsets of works and/or authors that are hard to nail down with traditional literary methods due to their size and the lack of outspoken thematic congruence. That bestselling crime fiction deals with police work and murder investigations is a fact by definition. But that literary fiction seems to have an almost equally close relation to nature and weather, for instance, is not obvious at all. In fact, the topic modelling approach enables a thematic view of literary fiction *as a category that shares certain thematic attributes*. This is not a minor claim. In some contexts, it might even be a radical one, as it lays bare the recurring tropes of the literary realm, almost as if literary fiction was a genre just as much as crime fiction is.

With a sociological book-trade perspective on literature, such an approach is unproblematic. In contemporary publishing, literary fiction *is* in most ways a genre similar to crime fiction.[69] What we are arguing here is that the same applies also on a textual level: contemporary Swedish literary fiction seems to deal with themes of nature, existential reflection, and intellectual development in more or less the same way as contemporary Swedish crime fiction concerns murders and police work. The big difference, of course, lies in how this is communicated to readers: where the genre conventions of crime fiction (and most other forms of genre fiction) are explicit, conventions for literary fiction are implicit—they are known by editors and readers as a taste, rather than understood as a genre. This claim needs to be developed further, in analyses that delve more deeply into the relationships among texts, book-trade mechanisms, and readers. But what we think we have managed to show here empirically is that (contemporary Swedish) literary fiction works as a genre not only in the marketplace, but also on the textual-thematic level.

The same goes for the two most prominent themes of the non-crime fiction bestsellers: food and economy. This outcome gives a clearer view of which themes bestsell-

ing books—apart from crime fiction—tend to contain. This subset consists of a blend of feelgood, chick lit, widely popular authors of literary fiction, and popularised literary fiction (e.g. through literary awards such as the August Prize). Still, they appear to unite in plots, settings, and scenes related to eating, drinking, money, work, and business.

Overall, two thematic poles of contemporary fiction that are not formally genre-bound seem to exist: the prosaic and the concrete pole on the one hand (common in popular fiction), and the poetic and the reflective one on the other hand (common in literary fiction). An assumption would be that the wider reading audience to a higher extent prefers to read about recognisable, contemporary, and often urban everyday life, whereas readers of literary fiction seem to prefer historical and retrospective novels set in the twentieth century, more often serene in nature and with many scenes taking place in rural or outdoor settings.[70]

These findings are worth exploring further. What distinguishes bestsellers from other books is a question as old as the book trade itself, but one that has never found a convincing answer.[71] Bestsellers can never be identified through textual analysis alone, since the commercial performance of books is a complex matter. Many factors, including textual qualities, book-trade structures, marketing, reception, genre, author, taste, reading habits, and word-of-mouth, are important. Yet, at least in a contemporary Swedish context, we think that our study undermines John Sutherland's famous claim that "what defines the bestseller is bestselling. Nothing else".[72] There are definitely patterns uniting bestselling fiction on a thematic level.

The final category of results consists of the surprising, or at first glance even contradictory, ones. In a way, these are of course the most interesting ones, as they seem to overturn common assumptions about Swedish literature. The first surprise is that topics on body parts, corporeality, intimacy, and violence are biased towards literary fiction rather than towards popular fiction. This outcome needs to be further investigated in order to be fully explained. The quantitative observations invite a closer reading of how sex and violence are depicted in contemporary popular fiction in contrast to literary fiction.

The second surprise is that topics relating to reading and writing do not exhibit any bias. On the contrary, they appear to be the most neutral ones in the whole experiment, both when it comes to the axis of prestige/popularity and to gender. While what looks like thematically close themes about education, intellectual development, and speech and language are clearly biased in literary fiction, as expected, the same does not apply to reading and writing. Apparently, readers of bestsellers and literary fiction both seem to enjoy reading about reading and writing to roughly the same extent. There are likely several explanations for this—self-identification as readers, fascination with the

writing process, nimbus of the literary realm—but it is a topic that deserves more in-depth investigation.

As is apparent, we call for a great deal of future work to follow up on our results. This is to be understood in relation to the exploratory vein of this article. A final such call is further development of computational methods for the analysis of themes in literary history.[73] But already, we think we have managed to show empirical results that provide new knowledge of contemporary Swedish fiction. Or to paraphrase Moretti: new conditions of knowledge of contemporary Swedish fiction.[74]

Tables

Table 1. The most evenly distributed topics between the subsets when k=20.

Topic	Top 10 words	"Theme"	Ratio: 1) Bs/Aug 2) BsNonCr/Aug 3) Male/Female	Outranking ratio: 1) Bs 2) BsNonCr 3) Male
12	man, kvinna, präst, bild, visa, heta, namn, [att] hälsa, dam, presentera [*man, woman, priest, picture/image, to show, to be called, name, greet, lady, to present*]	Meetings and presentations? (vague, mixed)	1) 0.96 2) 1.00 3) 1.05	1) 47% 2) 45% 3) 56%
17	skriva, läsa, bok, papper, tidning, brev, bild, dator, text, bläddra [*write, read, book, paper, newspaper/magazine, letter, picture/image, computer, text, browse*]	Reading and writing	1) 1.00 2) 0.94 3) 1.04	1) 49% 2) 46% 3) 55%
14	morgon, klocka, natt, kväll, sova, timme, vakna, vecka, eftermiddag, somna [*morning, clock, night, evening, sleep, hour, wake up, week, afternoon, fall asleep*]	Sleeping, waking up, time	1) 1.10 2) 1.05 3) 1.00	1) 61% 2) 51% 3) 48%
4	fråga, sak, förstå, berätta, mena, ord, svar, tala, ställa, sätt [*ask, thing, understand, tell, mean, word, answer, speak, put, way*]	Questions and answers, discussion	1) 1.02 2) 1.08 3) 1.10	1) 55% 2) 54% 3) 59%

Table 2. The most evenly distributed topics between the subsets when k=40.

Topic	Top 10 words	"Theme"	Ratio: 1) Bs/Aug 2) BsNonCr/Aug 3) Male/Female	Outranking ratio: 1) Bs 2) BsNonCr 3) Male
7	springa, skrika, slå, slita, falla, rusa, kasta, tappa, hoppa, fot [*run, scream, hit, tear/wrest, fall, rush, throw, drop/lose, jump, foot*]	Fights and scuffles	1) 0.97 2) 1.03 3) 1.01	1) 41% 2) 44% 3) 46%
22	skriva, läsa, bild, bok, tidning, brev, text, bläddra, dator, fotografi [*write, read, picture/image, book, newspaper/magazine, letter, text, browse, computer, photograph*]	Reading and writing	1) 0.96 2) 1.04 3) 1.03	1) 49% 2) 49% 3) 55%
27	man, kvinna, namn, heta, presentera, ålder, visa, dam, medelålder, efternamn [*man, woman, name, to be called, present, age, to show, lady, middle age, surname*]	Descriptions of people, introducing oneself or others	1) 1.05 2) 0.96 3) 1.09	1) 57% 2) 44% 3) 58%

Table 3. The most evenly distributed topics between the subsets when k=80.

Topic	Top 10 words	"Theme"	Ratio: 1) Bs/Aug 2) BsNonCr/Aug 3) Male/Female	Outranking ratio: 1) Bs 2) BsNonCr 3) Male
58	man, kvinna, ålder, skägg, medelålder, kön, karl, mustasch, yngling, utseende [*man, woman, age, beard, middle age, gender, man, moustache, young man, appearance*]	Descriptions of people	1) 1.09 2) 1.09 3) 0.99	1) 61% 2) 45% 3) 45%
20	papper, skriva, brev, dator, kamera, foto, penna, lapp, kuvert, dokument [*paper, write, letter, computer, camera, photo, pen, piece of paper, envelope, document*]	Reading and writing (documents, on computers, letters, notes, etc.)	1) 1.08 2) 0.92 3) 1.07	1) 56% 2) 39% 3) 52%

26	skrika, slå, slita, vakt, sparka, tag, gripa, tappa, golv, cell [*scream, hit, tear/wrest, guard, kick, grip, drop/lose, floor, cell*]	Fights, scuffles, and prison scenes	1) 1.08 2) 0.92 3) 1.07	1) 52% 2) 46% 3) 47%
38	sida, gata, plan, stanna, passera, vänster, höger, torg, butik, skylt [*side, street, plan, stop, pass, left, right, square, shop, sign*]	Crossing streets, travel and moving in cities	1) 1.00 2) 0.92 3) 1.12	1) 46% 2) 42% 3) 55%
16	folk, tycka, vän, verka, sätt, vis, bry, typ, idé, idiot [*people, think, friend, seem, way, way/manner, care, type, idea, idiot*]	(vague)	1) 1.00 2) 1.07 3) 0.98	1) 57% 2) 57% 3) 53%
24	stad, lämna, resa, vänta, sällskap, fara, humör, hälsa, dam, återvända [*city, leave, travel, wait, company, go, mood, greet, lady, return*]	Travel	1) 0.93 2) 1.00 3) 1.00	1) 46% 2) 57% 3) 55%
33	vecka, kväll, sommar, månad, helg, tillbringa, söndag, fredag, maj, jul [*week, evening, summer, month, weekend, spend, Sunday, Friday, May, Christmas*]	Vacation, holidays, spare time	1) 0.93 2) 1.00 3) 1.02	1) 60% 2) 58% 3) 52%
54	natt, sova, vakna, morgon, somna, dröm, kväll, drömma, sömn, väcka [*night, sleep, wake up, morning, fall asleep, dream, evening, to dream, awaken*]	Sleeping, going to bed, waking up	1) 0.92 2) 0.92 3) 0.93	1) 42% 2) 47% 3) 46%
55	springa, bit, hoppa, cykel, mur, varv, bänk, grind, väg, cykla [*run, bit/piece, jump, bicycle, wall, lap, bench, gate, way, to bike*]	Running, jumping, biking (vague)	1) 0.92 2) 0.92 3) 1.00	1) 41% 2) 49% 3) 45%

NOTES

1   For historical accounts on the relationship between popularity and prestige, see Andreas Huyssen, *After the Great Divide. Modernism, Mass Culture, Postmodernism*, Bloomington 1987; for computational experiments in this vein, see Ted Underwood, *Distant Horizons. Digital Evidence and Literary Change*, Chicago 2019, pp. 96–110.

2   Pierre Bourdieu, *The Rules of Art. Genesis and Structure of the Literary Field*, Cambridge 1996.

3   In this study, "theme" is used as "a type of literary content that is semantically unified and recurs with some degree of frequency or regularity throughout and across a corpus", to use

the definition of Matthew L. Jockers and David Mimno, "Significant themes in 19th-century literature", *Poetics*, vol. 41, 2013:6, pp. 750–769 (quote on p. 751, footnote 5). This is obviously a different definition than the one most familiar in literary scholarship. Instead of discerning the overarching ideas in literary works (often having to do with life, death, love, society, etc.), we are primarily interested in their manifest contents, i.e. what the narratives actually depict.

4   There are related projects, however, such as comparisons between the canon and the archive (Mark Algee-Hewitt et al., "Canon/Archive. Large-Scale Dynamics in the Literary Field", *Literary Lab Pamphlets*, vol. 11, January 2016), and between bestsellers and non-bestsellers (Jodie Archer & Matthew Jockers, *The Bestseller Code. Anatomy of the Block-Buster Novel*, New York 2016), as well as studies that map the relation between bestsellers and prestigious fiction, but not their content (J.D. Porter, "Popularity/Prestige", *Pamphlets of the Stanford Literary Lab*, vol. 17, September 2018; Underwood 2019, pp. 96–110). Also related is "The Riddle of Literary Quality", a project which computationally maps contemporary fiction in the Dutch book market (e.g. Andreas van Cranenburgh, "Cliché Expressions in Literary and Genre Novels", *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 2018, pp. 34–43; Corina Koolen, *Reading Beyond the Female. The Relationship Between Perception of Author Gender and Literary Quality*, Amsterdam 2018). In a Swedish context, there are also older studies comparing the predictability of the language (choice of words and sentence structure) in popular fiction and literary fiction, with predictable conclusions (e.g. Tor G. Hultman, "Text i bruk 4. 'den enkla form, som tusenden förstå'", in Ulf Teleman & Tor G. Hultman (eds), *Språket i bruk*, Lund 1974, pp. 277–296).

5   Franco Moretti, "Conjectures on World Literature", *New Left Review*, 2000:1, p. 57.

6   See Sarah Allison et al., "Quantitative Formalism. An Experiment", *Literary Lab Pamphlets*, 2011:1.

7   See e.g. Claire Squires, "Taste and/or Big Data? Post-Digital Editorial Selection", *Critical Quarterly*, 2017:3, pp. 24–38; Simone Murray, *The Digital Literary Sphere. Reading, Writing, and Selling Books in the Internet Era*, Baltimore 2018.

8   Andrew Goldstone, "The *Doxa* of Reading", *PMLA*, 2017:3, pp. 636–642.

9   Ted Underwood, "A Genealogy of Distant Reading", *Digital Humanities Quarterly*, 2017:2, URL: http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html, sec 32–38 (quote on sec 32); cf. Underwood 2019.

10   Katerine Bode, *A World of Fiction. Digital Collections and the Future of Literary History*, Ann Arbor 2018, p. 5.

11   Bode 2018, pp. 2–5, 17–35 (quote on p. 5).

12   In Swedish orig. "att systematiskt analysera relationerna mellan skönlitteratur och samhälle", Johan Svedjedal, "Det litteratursociologiska perspektivet. Om en forskningstradition och dess grundantaganden", *Litteratursociologi. Texter om litteratur och samhälle*, 2nd ed., Johan Svedjedal (ed.), Lund 2012, pp. 73–102 (quote on p. 78).

13   Svedjedal 2012, pp. 83–93.

14 We are not the first to highlight the bonds between distant reading and sociology of literature, see James F. English, "Everywhere and Nowhere. The Sociology of Literature After 'the Sociology of Literature'", *New Literary History*, vol. 41, 2010:2, pp. v–xxiii.

15 Sarah Allison, *Reductive Reading. A Syntax of Victorian Moralizing*, Baltimore 2018, pp. 1–35; cf. Franco Moretti, "'Operationalizing'. Or the Function of Measurement in Modern Literary Theory", *Literary Lab Pamphlets*, vol. 6, December 2013.

16 Allison 2018, p. 2.

17 Miha Kovač & Rüdiger Wischenbart, "Cultural Pluralism or Cultural Uniformity. Best-selling fiction books in Europe", *LOGOS*, 2009:1, pp. 249–261; Karl Berglund, *Deckarboomen under lupp. Statistiska perspektiv på svensk kriminallitteratur*, Uppsala 2012, pp. 77–79.

18 To single out crime fiction, we rely on "Deckarkatalogen", the annual bibliographies curated by Svenska Deckarakademin (The Swedish Crime Writer's Academy, URL: https://deckarakademin.org/hem/ [accessed 2019-06-26]). For a methodical account on the strengths of these bibliographies, see Berglund 2012, pp. 13–17.

19 Due to the nature of the bestseller charts and the way in which they are assembled, it is likely that our operationalisation of popular fiction misrepresents at least one segment of the Swedish book trade that is especially low in prestige, namely the romance paperbacks (and nowadays often e-books) published by Harlequin/HarperCollins. These still seem to be widely popular, but as they are often sold outside of the more common distribution models and do not seem to hinge on individual bestsellers, they are entirely lacking in our corpus.

20 From the beginning, the aim has been to draw attention to, and thus increase the sales of, the best Swedish books of the year. By giving booksellers a third of the votes, the idea was to ensure that the winning titles, elected just in time for the crucial Christmas sales, would be commercially viable. Poetry collections have thus often been shortlisted for the prize, legitimising the prize through association with cultural capital, but have in its first 30 years only won the prize twice. On the August Prize, see Jerry Määttä, "Prisade vare böckerna! Om Augustpriset", in *Böckernas tid. Svenska Förläggareföreningen och svensk bokmarknad sedan 1943*, ed. Johan Svedjedal, Stockholm 2018, pp. 577–651, and the web page of the August Prize: *https://www.augustpriset.se* [last accessed 2019-06-13].

21 The lines between the first two categories have occasionally been somewhat blurred, and nowadays autobiographies and memoirs are considered to belong in the category for adult fiction, and biographies in the category for adult non-fiction, while collections of essays can be considered in both categories, depending on their literariness. Of the 58 titles in the corpus (see below), fewer than ten are explicitly autobiographical, although many more clearly build on the personal lives and experiences of the authors, depicting protagonists that are sometimes barely masked versions of the authors.

22 The exception is Steve Sem-Sandberg's *De fattiga i Łódź*, which was released as a Swedish e-book in September 2019, some months after our analyses were conducted.

23 The corpus thus includes works by authors such as Sigrid Combüchen, Per Olov Enquist, Monika Fagerholm, Torgny Lindgren, Carl-Henning Wijkmark, and Klas Östergren, but

also by younger writers such as Johannes Anyuru, Jonas Hassen Khemiri, and Sara Stridsberg.

24 Such examples are Johannes Anyuru's *De kommer att drunkna i sina mödrars tårar* (2017), Carl-Michael Edenborg's *Alkemistens dotter* (2014), and John Ajvide Lindqvist's *Rörelsen* (2015), with their clear influences from science fiction, fantasy, and horror respectively.

25 James F. English, *The Economy of Prestige: Prizes, Awards, and the Circulation of Cultural Value*, Cambridge, MA. 2005, pp. 3–11, 51, 207–208, 234 et passim.

26 Robert Escarpit, *Sociologie de la littérature* (1958), quatrième édition mise à jour, Paris 1968, pp. 90–97.

27 Hans Hertel, *500.000 £ er prisen. Bogen i mediesymbiosens tid*, The Adam Helms Lecture 1995, Stockholm 1996, pp. 32–37.

28 The 9 winning titles on the bestseller charts are Bengt Ohlsson's *Gregorius* (2004), Monika Fagerholm's *Den amerikanska flickan* (2005), Susanna Alakoski's *Svinalängorna* (2006), Carl-Henning Wijkmark's *Stundande natten* (2007), Sigrid Combüchen's *Spill* (2010), Lena Andersson's *Egenmäktigt förfarande* (2013), Kristina Sandberg's *Liv till varje pris* (2014), Jonas Hassen Khemiri's *Allt jag inte minns* (2015), and Johannes Anyuru's *De kommer att drunkna i sina mödrars tårar* (2017). The other 5 shortlisted titles on the bestseller charts are Per Olov Enquist's *Boken om Blanche och Marie* (2004), Klas Östergren's *Gangsters* (2005), Jonas Hassen Khemiri's *Montecore* (2006), Torgny Lindgren's *Norrlands Akvavit* (2007), and Åsa Linderborg's *Mig äger ingen* (2007).

29 Cf. Hertel 1995; Claire Squires, *Marketing Literature. The Making of Contemporary Writing in Britain*, Basingstoke 2007, pp. 70–104.

30 For a thorough investigation on the usability of the paragraph in large-scale literary modelling, see Mark Algee-Hewitt et al., "On Paragraphs. Scale, Themes, and Narrative Form", *Stanford Literary Lab Pamphlets*, vol. 10, 2015.

31 See, for instance, David M. Blei, Andrew Y. Ng, & Michael I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, 3, 2003, pp. 993–1022, Mark Steyvers & Tom Griffiths, "Probabilistic topic models", in T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (eds), *Latent Semantic Analysis. A Road to Meaning*, Mahwah, NJ. 2007, pp. 427–448.

32 Mats Dahllöf & Karl Berglund, "Faces, Fights, and Families. Topic Modeling and Gendered Themes in Two Corpora of Swedish Prose Fiction", in Costanza Navaretta et al. (eds), *DHN 2019 Copenhagen, Proceedings of 4th Conference of The Association Digital Humanities in the Nordic Countries*, 2019.

33 We used the efselab tagger. Robert Östling, "Part of Speech Tagging. Shallow or Deep Learning?", *North European Journal of Language Technology* 5, 2018.

34 I.e. treated as *stop words*.

35 Andrew Kachites McCallum, "MALLET. A Machine Learning for Language Toolkit", http://mallet.cs.umass.edu, 2002–2018. [accessed 2019-11-25]

36 Dahllöf & Berglund 2019.

37 This kind of ratio is called *risk ratio* (in epidemiology).

38 As explained in Dahllöf & Berglund 2019.

39 http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-397762

40 Our experiment is based on nouns and verbs in the novels. These part-of-speech tags are of course noted in the output but are here, except in ambiguous cases, removed for readability.

41 These include tracking and crime scene investigations (k=40, topic 33), time and meetings (k=80, topic 4), shootings (k=80, topic 79), perpetrators, victims, and evidence (K=160, topic 95), and interrogations and criminal procedure (k=160, topic 142). For details on these topics, see the supplementary material.

42 For instance, lighting and weather (k=40, topic 10), education and intellectual development (k=40, topic 4), speech and language (k=80, topic 67), and smoking and fire (k=160, topic 156), along with several more fine-grained topics relating to outdoor settings, such as maritime milieus (21), weather, celestial phenomena, light and colour, and/or forests (46, 77, 52). For details, see the supplementary material: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-397762.

43 The same goes for titles such as Per Gunnar Evander's *I min ungdom speglade jag mig ofta* (2005), Daniel Sjölin's *Världens sista roman* (2007), Eva Adolfsson's *En liten historia* (2009), and corpus-overlapping titles such as Per Olov Enquist's *Boken om Blanche och Marie* (2004), Carl-Henning Wijkmark's *Stundande natten* (2007), and Lena Andersson's *Egenmäktigt förfarande* (2013).

44 Of the ten works most closely associated with this topic, seven were written by Guillou.

45 The same would seem to apply for translated bestsellers in Sweden, as there are very few international writers of thrillers and espionage in the vein of Robert Ludlum or John Le Carré on the Swedish bestseller charts, one of the few possible exceptions being Dan Brown.

46 These authors have written novels that deal with the upper echelons of eighteenth- and nineteenth-century Sweden (Persson), twentieth-century history (Enquist), or riots (Gassilewski), and genocides (Rosenberg, Fröberg Idling).

47 Our bestseller corpus does include a large quantity of crime fiction, but almost no romance (and no erotica), since such genres do not often end up on print bestseller charts in Sweden and are often consumed in other ways (as e-books, audiobooks, etc.).

48 E.g. "smeka", "svett", "andetag", "andas", "kyssa", "blunda" [*caress, sweat, breath, breathe, kiss, close one's eyes*] vs. "blod", "smärta", "sår", "gråta", "skära", "kräkas" [*blood, pain, wound, cry, cut, vomit*].

49 See e.g. Annika Hällsten, "Brutalt, äckligt och blodigt", *Hufvudstadsbladet*, 2017-03-18; Kerstin Bergman, "Mer mord och mörker för Kepler-frälsta", *Smålandsposten*, 2018-10-29.

50 Cf. Karl Berglund, *Död och dagishämtningar. En kvantitativ analys av det tidiga 2000-talets svenska kriminallitteratur*, Uppsala 2017, p 94. Many other scholars have stated that Swedish crime fiction is occupied with depictions of severe violence, especially in relation to sexual abuse and aberrations. See e.g. Katarina Gregersdotter, "Mördaren i garderoben. Bilden av den omanlige mördaren i *Steget efter* och *Nattsystern*", *Tidskrift för litteraturvetenskap*, 2009:2, pp. 53–61; Anna Westerståhl Stenport & Cecilia Ovesdotter Alm, "Corporations, Crime, and Gender Construction in Stieg Larsson's *The Girl with the Dragon*

Tattoo. Exploring Twenty-First Century Neoliberalism in Swedish Culture", *Scandinavian Studies*, 2009:2, pp. 157–178; *Men Who Hate Women and Women Who Kick their Asses*, Donna King & Carrie Lee Smith (eds), Nashville 2012; *Rape in Stieg Larsson's Millennium Trilogy and Beyond. Contemporary Scandinavian and Anglophone Crime Fiction*, Berit Åström, Katarina Gregersdotter & Tanya Horeck (eds), Basingstoke 2013.

51 Cf. Karl Berglund, *Mordförpackningar. Omslag, titlar och kringmaterial till svenska pocketdeckare 1998–2011*, Uppsala 2016, pp. 40–42, 157–162.

52 Cf. Karl Berglund, *Snille och smak och mord. Om genreproblematik och Nobelpristematik i Karin Alvtegens Skugga*, Stockholm 2010, pp. 30–52.

53 See Underwood 2019, pp. 19–20.

54 Cf. Underwood 2019, pp. 19–20.

55 No works of crime fiction were shortlisted for the August Prize between 2004 and 2017, while 14 titles shortlisted for the prize ended up on the bestseller charts. Of the 57 non-crime fiction bestsellers and the 58 titles shortlisted for the August Prize that are used in these comparisons, 14 works are present in both sub-corpora; roughly a quarter (ca. 25%) of them are thus identical.

56 But not all of them, for instance not themes connected to police work, which are so slightly biased that they lean towards the more or less neutral segment of topics, as well as for "state affairs, people with power, jurisdiction" (k=40, topic 20).

57 Topics on "Money and economy": k=40, topic 37; k=80, topic 50; k=160, topics 119, 9. Topics on "Telephoning": k=40, topic 35; k=80, topic 68; k=160, topic 2.

58 The strange "n" can probably be explained by the isolated "n" in words like "TV:n" [*the TV*], which the tagger did not handle properly.

59 Some are the same as in the earlier comparisons with bestsellers at large—"existential reflection and strong emotions" (k=20, topic 1), "body parts, corporeality, intimacy, and violence" (k=20, topic 3), and "family" (k=160, topic 111)—some are new, such as "reading and writing" (k=20, topic 17) and "doctors, hospitals, and illness" (k=40, topic 40; k=80, topic 76).

60 To some extent, these differences reflect the above-noted differences of plot versus setting, but not as distinctly as for bestsellers at large. Apart from "telephone calls", topics reflecting plot are less biased in the non-crime fiction sub-corpus.

61 A difference from the food topics, however, is that the top levels for individual works are lower here. Instead, the topics related to "money and economy" appear as the most biased due to there being very few non-crime fiction bestsellers at the bottom, dominated by prestigious fiction and crime fiction. The latter is interesting, since there are simultaneously several crime-fiction titles in the absolute top segment—all Jens Lapidus's novels, most notably (and not surprising, since all his novels since the debut *Snabba cash* have dealt with organised crime, financial crime, and money laundering), but also crime fiction by writers such as Jan Guillou, Stieg Larsson, and Leif GW Persson. Thus, "money and economy" seems to be more stably predominant among non-crime fiction bestsellers than in bestselling crime fiction, with higher lows, but also lower highs.

62 Topic 12 seems to be a mixture of themes merged in the modelling process due to the low k-value, but with a slight tendency towards presentations of characters and meetings between people (and an ecclesiastical setting); topic 4 appears to relate mostly to dialogue, questions and answers, and discussions.

63 The slight over-representation that can be noted in the bestseller corpus might be explained by their general focus on work, a theme that is strongly related to scheduling matters, but perhaps also to sleep (or lack of sleep, a common trope among the investigators in contemporary Swedish crime fiction).

64 Furthermore, some other themes with an even distribution in the subset pop up. Two are recurring: "descriptions of people" (k=40, topic 27; k=80, topic 58) and "fights and scuffles" (k=40, topic 7; k=80, topic 26) (see table 2–3). While the former is not surprising, the latter is, at least to some extent. In an earlier paper, we discussed topics related to fights in contemporary bestsellers. We saw how a larger number of topics gave us more fine-grained categories of gun fights, fist fights, and more serious injuries (Dahllöf & Berglund 2019, pp. 100–102). However, we more or less assumed that this theme was specific to bestsellers and not to contemporary fiction in general. The results derived here indicate that at least "fights and scuffle" seems to be equally present in literary fiction, whereas "gun fights" shows a clear bias towards bestsellers (k=80, topic 79, bs/aug ratio: 1.60).

65 Most recently by Alba Mogensen and Amanda Svensson (Mogensen, "Författarna som väljer bort handlingen", *Göteborgs-Posten*, 2019-07-07; Svensson, "Låt inte storytelling döda den stora berättelsen", *Expressen*, 2019-07-18), but perhaps most famously in the debate following the "Manifesto for a new literary decade", published by seven renowned writers in Sweden's largest newspaper in 2009 (Susanne Axmacher, Jesper Högström, Sven Olov Karlsson et al., "Manifest för ett nytt litterärt decennium", *Dagens Nyheter*, 2009-08-22).

66 This category of findings resembles Ted Underwood's argument that results derived from distant readings might not always provide causal explanation in themselves. Instead, they can be used to support earlier, qualitative claims: "Although it sounds paradoxical, we should not be surprised to find that quantitative analysis sometimes reveals unexpected patterns—for which we already have several good explanations handy" (Underwood 2019, p. 139).

67 The prestige- and gender-neutral constitution of themes about sleeping, time and scheduling, and descriptions of people's attributes could be added.

68 As Corina Koolen notes, however, it is important not to draw too big conclusions from such an outcome. There are always individual authors who break the patterns, and most topics are not gender-biased at all. (Koolen 2018, pp. 158–160.)

69 See e.g. Squires 2007, pp. 70–104.

70 Such an argument resembles Karl Berglund's explanation of the emphasis given to everyday life in contemporary Swedish crime fiction (Berglund 2017, pp. 127–128).

71 Although many have tried, most recently Archer & Jockers 2016.

72 John Sutherland, *Bestsellers. A Very Short Introduction*, Oxford 2007, p. 22.

73 One way to do this would be to complement the unsupervised topic modelling with a supervised approach. It would be possible, for instance, to check the presence of themes by

using hand-crafted dictionaries of terms which are strongly and unambiguously associated with them.

74 Cf. Moretti 2000, p. 57.

ABSTRACT

Karl Berglund, Department of Literature, Uppsala University
Mats Dahllöf, Department of Linguistics and Philology, Uppsala University
Jerry Määttä, Department of Literature, Uppsala University

Apples and Oranges? Large-Scale Thematic Comparisons of Contemporary Swedish Popular and Literary Fiction

The aim of this article is to compare thematic trends in contemporary Swedish bestselling and literary fiction with the help of a computational method—topic modelling—which extracts content themes based on statistical patterns of word usage. This procedure allows us to identify trends and patterns that are not easily discovered through manual reading. We track topics in two subsets of Swedish fiction from the period 2004–2017: 1) prose fiction on the Swedish bestseller charts, and 2) prose fiction shortlisted for the August Prize (arguably the most prestigious Swedish literary prize). The results confirm several assumptions about contemporary popular and literary fiction, such as more plot-focused themes in popular fiction and themes more connected to settings in literary fiction. But the outcomes also provide new, and more surprising knowledge, such as food and economy being the most biased themes among the non-crime fiction bestsellers, whereas themes concerning nature are most biased in the literary realm. Moreover, themes relating to sex, intimacy, and violence are biased towards literary fiction rather than popular fiction. In the light of our findings, we argue that both popular fiction and literary fiction seem to be characterised by certain thematic attributes that make it relevant to discuss them as genres also on a textual-thematic level.

Keywords: Bestsellers, The August Prize, Swedish literary fiction, distant reading, topic modelling, popular fiction