

More Social, Less Religious: Trends of Hardcover Fiction Titles on the New York Times Bestseller List 2000-2020

Bei der Wieden, Gesa

g.beiderwieden@stud.uni-goettingen.de
University of Göttingen, Germany

Bathurst, Taylor

taylor.bathurst@stud.uni-goettingen.de
University of Göttingen, Germany

Haider, Thomas Nikolaus

thomas.haider@uni-goettingen.de
University of Göttingen, Germany; Max Planck Institute for
Empirical Aesthetics, Frankfurt

the economic and cultural environment. He found that with a significantly growing book market, titles became much shorter, and also that titles from the beginning of the nineteenth-century reflect nineteenth-century ethics. Thus, titling practices follow the book market, but also presumably the cultural and aesthetic preferences of readers.

Corpus

To study the trends in titling practices, we compiled a corpus of all titles that have been on the New York Times (NYT) best-seller list for hardcover fiction, between 2000 and 2020, which is based on sales data from all over the United States (New York Times Best-Seller Lists Staff, 2020). For any given year, a title is included if it entered the list in that year, regardless of how long it stayed. Although the online published list carries more information (weekly rankings), we start with a dataset from *data.world* that already includes yearly lists from February 2011 until June 2018 (cf. Tauberg, 2018), and augment it through manual collection to cover the years from 2000 to 2020. We also constrain the data to hardcover fiction, albeit the NYT publishes genre-specific rankings as well. We leave a more fine-grained analysis to future work. Overall, we retrieve over 4000 titles in total. See Table 1 for exact numbers of titles and tokens per year. The data can be found at <https://github.com/tnhaider/nyt-fiction-bestseller-list>

Abstract

We examine changes in book titles of the New York Times Best-seller hardcover fiction lists from 2000 to 2020 regarding thematic, grammatical, and psychological categories. We find that social categories (including personal pronouns) have been increasingly used, while religiously connotated titles have been declining.

Introduction

The title — “half sign, half ad” (Moretti 2009) — is a reader’s entry point to a novel. Research on how people select books in online library catalogs underlines that title names play an important role in attracting readers (Mikkonen and Vakkari, 2016; Pöntinen and Vakkari, 2013). Readers select books with titles that appeal to their personal life experiences (Mikkonen and Vakkari, 2016), such as through references to gossip, morality, pleasures, and mental processes (Barnes / Black 2022). However, while studies on the aesthetics (and success of) books typically do not discuss titles (Ashok et al., 2013; Yucesoy et al., 2018; Wang et al., 2019), research on fiction titles focuses on historical corpora (Patras et al., 2021; Moretti, 2009; Benard, 1995), or on other domains, such as a corpus of news headlines (Oberländer et al., 2020). Several studies on titles of scientific articles also exist (cf. Cheng et al., 2012; Sahragard and Meihami, 2016; Hartley, 2007).

When analyzing titles, we should consider which functions they fulfill, and which cultural significance they bear. According to Genette (1998), a title carries the following functions: a) identifying the text (of a book), b) describing its content, c) holding connotative value, and d) attracting readers. Studying British novels between 1740 and 1850, Moretti (2009) highlighted the relationship between the length, syntax, and semantics of titles and changes in

Year	# Titles	# Tokens
2000	137	398
2001	130	404
2002	139	413
2003	142	417
2004	148	427
2005	156	417
2006	176	464
2007	183	508
2008	193	509
2009	199	520
2010	196	519
2011	220	599
2012	255	730
2013	333	903
2014	363	1039
2015	337	918
2016	359	977
2017	252	712
2018	229	641
2019	186	547
2020	165	508

Table 1: Number of titles and tokens per year.

Experiments

To measure the change in titling practices, we use the Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2015), a dictionary that maps words to grammatical, psychological, and thematic categories. We treat each year as a bag-of-words, for which we extract all associated categories via the python package *liwc* (Brown, 2012). We then calculate the relative frequency of each category for a given year. We also calculate the length of titles and count unigram frequencies (excluding a set of stopwords).

Based on the relative category frequencies per year for each category, we estimate their trend by calculating Spearman rho with the year range. This shows us whether, and how strongly, the trend for a category increases or decreases. Furthermore, the p-value shows us whether the trend is significant or not. We choose $p=0.05$ as the threshold. See Table 2 for the significant categories.

LIWC category	Correlation coefficient	p-value	Example words
Total function words	0.4649	0.0337	it, to, no, very
Total pronouns	0.6535	0.0013	I, them, itself
Personal pronouns	0.6208	0.0027	I, them, her
1st person plural	0.6866	0.0006	we, us, our
1st person singular	0.4558	0.0378	I, me, mine
Impersonal pronouns	0.5117	0.0177	it, it's, those
Social processes	0.7247	0.0002	people, talk, they
Family	0.4819	0.0269	daughter, dad, wife
Cognitive processes	0.4896	0.0243	and, know, secret
Certainty	0.4961	0.0222	always, never
Religion	-0.7104	0.0003	christmas, devil, faith
Assent	0.4561	0.0377	absolutely, yes
Filler	0.4541	0.0386	like

Table 2: Result of Trend Analysis with Spearman rho on LIWC categories. Numbers rounded to four (five) decimal places.

Our analysis shows that titles of bestselling fiction books increasingly contain features that previous research has shown to be appealing to readers. People like titles alluding to mental processes and gossip (Barnes / Black 2022). Topic modelling on the contents of bestselling books has shown that human closeness and human connection are topics that are characteristic of bestsellers (Archer and Jockers, 2016: 67). Furthermore, people like titles they can relate personal experiences with (Mikkonen and Vakkari, 2016).

We have found increases in mentions of family terms (presumably reliable for most). Generally speaking, social categories dominate the significant upward trends in our data (personal pronouns, family, social processes), meaning that words related to social categories like *people*, *talk* and *we*, are relatively more present in later years (Figure 1), rising from a relative share of 5% of all tokens in 2000 to a share of 11% in 2019. Close reading of titles referring to social categories indicates that this increase is attributable to interpersonal matters (e.g., romance, coming of age), rather than social issues, particularly seen in titles containing the pronouns *we* and *you*.

Besides social categories, our results show that words describing cognitive processes (particularly those expressing certainty) and terms referring to spoken categories (assent and fillers) increased over the years.

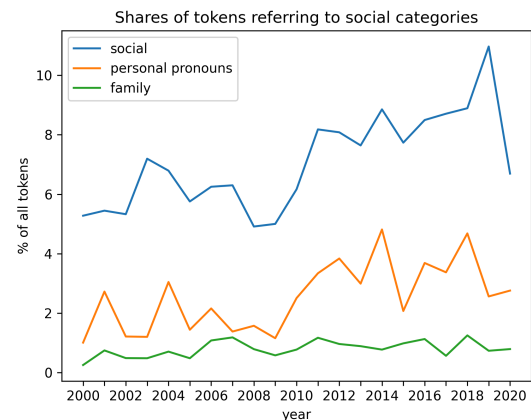


Figure 1: Trend for rising categories ‘social processes’, ‘personal pronouns’ and ‘family’.

However, the share of religious terms has decreased significantly since 2000 (apart from a sudden peak in 2008) (Figure 2). A majority of years feature *christmas* titles, and especially in the earlier years (e.g., 2008) there are additionally a number of references to e.g., devils and angels, but frequently in a metaphorical

sense without clear religious content, rather alluding to *good* and *evil*, or a specific attitude (as in the James Bond novel *Devil may care*, or *Seducing an Angel* by Mary Balogh). Furthermore, research on American religiosity declining (Voas and Chaves, 2018) may help contextualize the decrease of terms referring to religion.

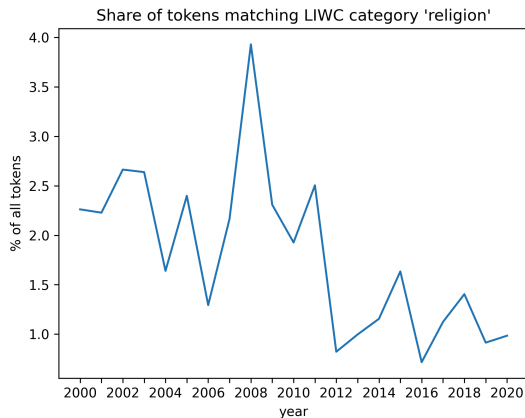


Figure 2: Trend for falling category 'religion'.

Finally, our method focused on trend analysis, but we did not study terms (or categories) that have remained popular in detail. An analysis of the most frequent words over the years reveals that the words *girl(s)* and *death* have maintained a high share, similar to Haider (2019), where it is shown that these topics have stayed frequent in German poetry since 1600. However, this could be a genre effect here, since mystery and thriller books are among the bestselling fiction genres (Yucesoy et al., 2018). Furthermore, since other research found changes in the length of titles (Moretti, 2009; Tauberg, 2019), we examined the length of titles, as shown in Figure 3, finding no substantial change, where most titles are between two and three words long.

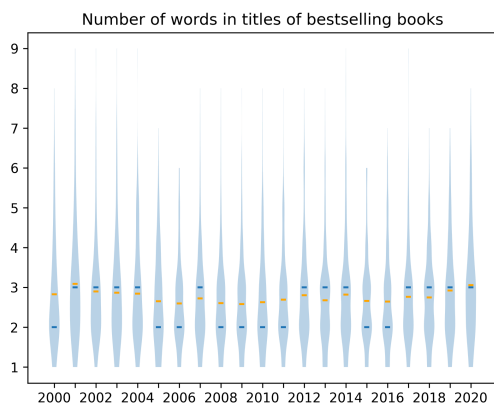


Figure 3: Violin plot depicting the distribution of title lengths per year. The width corresponds to the estimated density of data points. Averages indicated in orange, medians in blue.

Conclusion & Future Work

Our analysis highlights trends in psychological, thematic, and structural features of bestselling fiction titles. We have shown that titles containing words related to social and cognitive categories have been increasingly adopted over the last two decades. However, references to religious terms have been on the decline, and that many religious references are regarding metaphorical uses of moral categories (good, evil).

Future research might want to expand the dataset (weekly lists, or expanding the time frame), to study finer trends. Furthermore, studying syntactic features and group-based differences (bestselling vs. non-bestselling titles, genres) could yield interesting results. Finally, identifying confounds was beyond the scope of this paper, for example whether (and how much) the length of titles has an influence on certain trends (e.g., social titles becoming longer).

Bibliography

- Archer, Jodie / Jockers, Matthew L.** (2016): *The Bestseller Code. Anatomy of the Blockbuster Novel*. New York: St. Martin's Press.
- Ashok, Vikas G. / Feng, Song / Choi, Yejin** (2013): "Success with Style: Using Writing Style to Predict the Success of Novels", in: Association for Computational Linguistics (ed.): *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, 2013: 1753–1764.
- Barnes, Jennifer L. / Black, Jessica E.** (2022): "What's in a name? Book title salience and the psychology of fiction", in: *Psychology of Aesthetics, Creativity, and the Arts* 16, 2: 290–300.
- Benard, Michel** (1995): "Àjuste titre: a Lexicometric Approach to the Study of Titles", in: *Literary and Linguistic Computing* 10, 2: 135–141.
- Brown, Christopher** (2012): *Linguistic Inquiry and Word Count (LIWC) analyzer*. Python package <https://github.com/chbrown/liwc-python> [04.05.2023].
- Cheng, Stephanie W. / Kuo, Chih-Wei / Kuo, Chih-Hua** (2012): "Research article titles in applied linguistics", in: *Journal of Academic Language and Learning* 6, 1: A1–A14.
- Genette, Gérard** (1988): "Structure and Functions of the Title in Literature" (Trans.: Crampé, Bernard), in: *Critical Inquiry* 14, 4: 692–720.
- Haider, Thomas** (2019): "Diachronic Topics in New High German Poetry", in: Alliance of Digital Humanities Organizations (ed.): *Proceedings of the Digital Humanities Conference 2019*, Utrecht, 2019: arXiv:1909.11189.
- Hartley, James** (2007): "Planning that title: Practices and preferences for titles with colons in academic articles", in: *Library & Information Science Research* 29, 4: 553–568.
- Mikkonen, Anna / Vakkari, Pertti** (2016): "Readers' interest criteria in fiction book search in library catalogs", in: *Journal of Documentation* 72, 4: 696–715.
- Moretti, Franco** (2009): "Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740–1850)", in: *Critical Inquiry* 36, 1: 134–158.
- New York Times Best-Seller Lists Staff** (2000–2020): "Hardcover Fiction", in: *The New York Times Book Review* <https://www.nytimes.com/books/best-sellers/hardcover-fiction/> [04.05.2023].
- Oberländer, Laura A. M. / Kim, Evgeny / Klinger, Roman** (2020): "GoodNewsEveryone: A Corpus of News Headlines Annotated with Emotions, Semantic Roles, and Reader Perception",

in: European Language Resources Association (ed.): *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Marseille, 2020: 1554–1566.

Patras, Roxana / Odebrecht, Carolin / Galleron, Ioana / Arias, Rosario / Herrmann, Berenike J. / Krstev, Cvetana / Poiniz, Katja M. / Yesypenko, Dmytro (2021): "Thresholds to the 'Great Unread': Titling Practices in Eleven ELTeC Collections", in : *Interférences littéraires/Littéraire interferences* 25: 163-187.

Pennebaker, James W. / Boyd, Ryan L. / Jordan, Kayla / Blackburn, Kate (2015): *The development and psychometric properties of LIWC2015*. Austin: University of Texas.

Pöntinen, Janna / Vakkari, Pertti (2013): "Selecting Fiction in Library Catalogs: A Gaze Tracking Study", in: *International Conference on Theory and Practice of Digital Libraries*, Valletta, 2013: 72-83.

Sahragard, Rahman / Meihami, Hussein (2016): "A diachronic study on the information provided by the research titles of applied linguistics journals", in: *Scientometrics* 108: 1315–1331.

Tauberg, Michael (2018): "New York Times Bestsellers from 2011 to 2018", in : *data.world*. <https://data.world/typhon/new-york-times-bestsellers-from-2011-to-2018> [04.05.2023].

Tauberg, Michael (2019): "Are Book Titles Getting Longer?", in: *Towards Data Science*. <https://towardsdatascience.com/book-titles-are-getting-longer-1c341fbd4829> [04.05.2023].

Voas, David / Chaves, Mark (2018): "Even intense religiosity is declining in the United States: Comment", in: *Sociological Science* 5: 694-710.

Wang, Xindi / Yucesoy, Burcu / Varol, Onur / Eliassi-Rad, Tina / Barabasi, Albert-László (2019): "Success in books: predicting book sales before publication", in: *EPJ Data Science* 8: 1-20.

Yucesoy, Burcu / Wang, Xindi / Huang, Junming / Barabási, Albert-László (2018): "Success in books: a big data approach to bestsellers", in: *EPJ Data Science* 7: 1-25.