ImpFic - Impact of Fiction

Proposal for Accelerating Scientific Discovery 2020 / Netherlands eScience Center
Applicants: Dr. Peter Boot, Dr. Marijn Koolen, Drs. Joris van Zundert / Huygens ING

Abstract

In this project we ask which properties of fictional texts have an impact on readers. The impact types we look at include affective responses to narrative and style as well as reflection. We distinguish different groups of readers and require the textual properties to be meaningful to literary researchers and readers. Earlier research often handled all readers as essentially similar and targeted a single measure of success (popularity or sales). It also tried to predict success based on features that are hard to interpret from a literary point of view (such as word frequencies). We will use a large corpus of recent Dutch novels, a large corpus of online book reviews and a large collection of book lists created by users on book-oriented social media sites. In the reviews, we measure different types of impact. Based on the book lists, we cluster readers by their preferred type of reading. For the texts, we define new metrics for key textual properties that (we hypothesize) are partly responsible for the impact a book has on its readers. These metrics will include parameters referring to the novel's narrative, writing style and mood.

Background and Research Question

Our **objective** is to investigate how reading fiction influences readers. We define, compute and assess textual properties in Dutch fiction that are both meaningful for readers and literary researchers and computable from the text. Reading of fiction, specifically literature, is widely believed to be good for well-being, citizenship and society (Hovinga 2019). Yet, quantitative research on exactly how reading influences readers remains scarce.

The fundamental **research questions** are:

1. How does reading fiction affect or impact readers?
2. Which textual properties of books contribute to this impact?
3. How does this impact depend on the reader's reading preferences?

Literature is to some extent 'a market of symbolic goods' (Bourdieu 1985) and its effects also depend on external factors, such as prestige (De Nooy 2002). However, there exists ample proof that the book text is also an important factor determining the effect of books (e.g. Gavaler & Johnson 2019). Our **approach** combines book texts, online reviews, online book lists and book sales and lending numbers. Our goal is to establish how textual characteristics determine the impact of fiction that readers report in their book reviews as well as sales and lending figures. The impact factors we distinguish are response to narrative, aesthetic response and reflection (Koopman & Hakemulder 2015) using the approach of Boot and Koolen (2020). We deduce reading preference by clustering readers based on the book lists that readers create on online reading platforms. The biggest challenge, however, is in establishing and operationalizing the textual variables. We address this in section 6b.

Unavoidably, we have to choose from the many relevant textual aspects that we could study. We make our choice based on variety in content and expected difficulty. Our **use cases** are:

1. two important **overall textual properties**: mood or emotional tone (Archer & Jockers 2016) and topic;
2. a number of **stylistic properties**: concreteness, densities of action, dialogue, reflection and description in the text, complexity, unexpectedness (deviation from 'normal' style; Miall & Kuiken 1999) and some reader-level stylistic characteristics (dramatic, elaborate, simple and thoughtful; taken from Saricks 2005);
3. two properties of the **narrative**: attractiveness of characters (Maslej 2019) and narrative pace.

We believe this project will have considerable **impact**, for a number of reasons:

1. Up to now, in computational research, book texts and reviews have been studied separately. We bring together these distinct data types.
2. The NLP tools developed in this project can be reused and serve as building blocks for further research on narrative and storytelling and on recommender systems.
3. We provide a computational underpinning for qualitative, experiential concepts about literature.

State of the art: There are three areas of related research. The first area concerns computational research on how popularity and ratings of novels correspond to textual features such as n-grams and part-of-speech-tags (nouns, verbs, etc.) (e.g. Ashok, Feng & Choi 2013, Van Cranenburg & Bod 2017, Louwerse et al. 2008). These features are hard to interpret from a literary or psychological perspective.

The second area investigates readers, be it in classrooms (e.g. Wilhelm 2016), through surveys (e.g. Koolen et al. 2020), or in the laboratory (e.g. Miall 2006). In this last approach, the impact of literary characteristics is measured directly in readers. It is often based on the notion of stylistic deviance leading to foregrounding of meaning in the mind of the reader. Studied features include literary vs. non-literary text and fiction vs. non-fiction (Kidd & Castano 2013), episode structure (Miall, 2004), foregrounded textual features (Koopman 2016) or presence of a narrator (Andringa 1996). These methods by themselves provide no way to computationally measure these features in texts.

The third area investigates reader impact using book reviews, (Rebora et al 2019 provide an overview). This is predominantly focused on responses to individual books (Gutjahr 2002; Harris 2019), although Driscoll and Rehberg Sedo (2018) have more systematically investigated reader experiences in reviews. The approach of Lendvai et al. (2020) is closest to our inquiry, measuring narrative absorption in stories based on a manually annotated set of sentences from online reviews. These studies demonstrate that reviews are a useful source to study reader impact, but do not make the link to textual features of novels.

Our project brings these three strands of research together to investigate the connection between textual features of novels and reader impact.

Technological Challenges

The bottleneck preventing further advances in this field is that there are many Natural Language Processing (NLP) resources available for low-level features -- part-of-speech-tags, named entities, sentiment -- but few for deriving features from narrative text at the level that researchers and readers care most about (plot, pace, mood, attractiveness of characters). We need to address this.

The technological challenge is thus to build an NLP and Machine Learning (ML) pipeline that takes as input a novel and a set of reviews of that novel and computes or learns the features expressing concepts from the domain of literary studies. Identifying or measuring different high-level features in novels requires different sequences and combinations of analyses from a range of existing NLP tools.

Our basic workflow is visualized in Figure 1 and consists of four parts. We combine three types of data (see Table 1): about readers and reading behaviour (1 in Figure 1), reading impact scores and reader judgements (2), and features extracted from book texts (3), and learn how book features contribute to reader judgements using statistical inference and ML (part 4). Relying on manual annotation of the novels with high-level concepts for machine learning is too time consuming on this scale and results in black boxes instead of features interpretable to humans. We explain the main components of the figure below and show how this might work for a number of specific research questions.

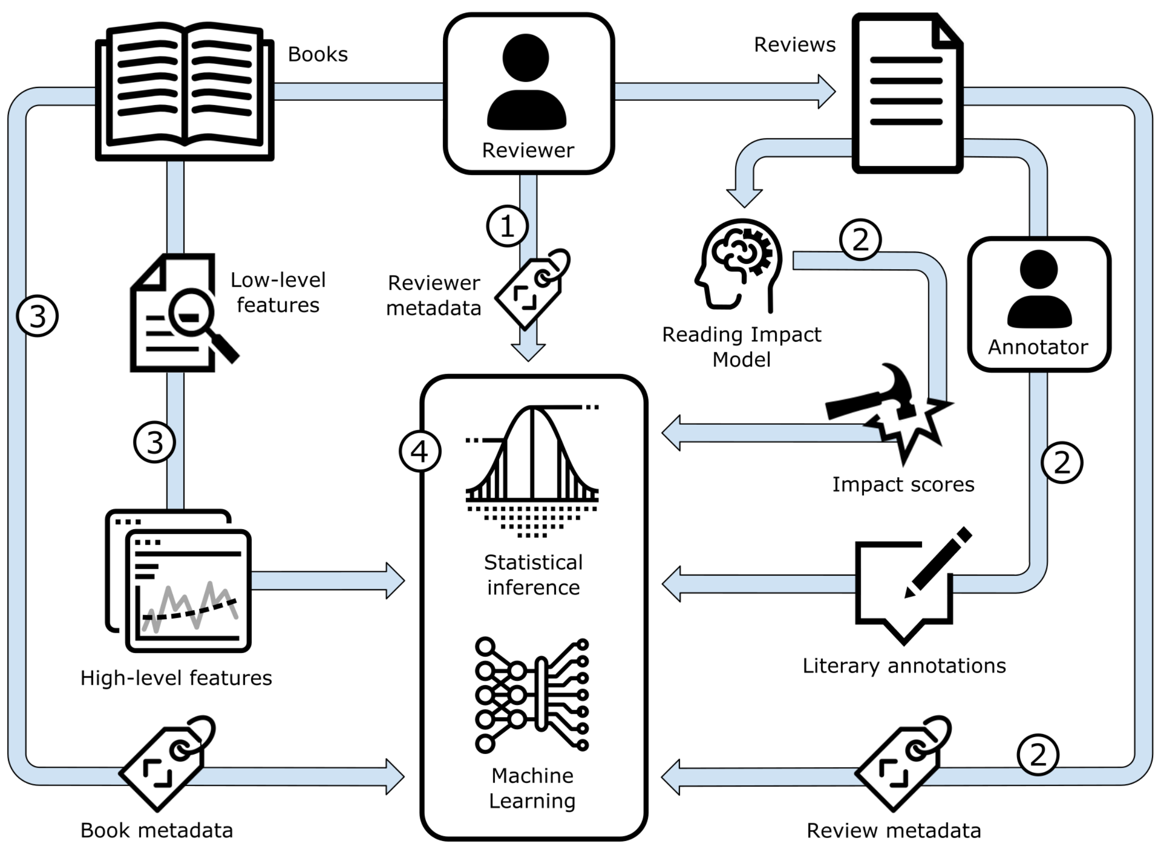

Fig. 1. High-level overview of the processes and data collections in our project.

Datasets	Description
Book corpus	A large number (tbd) of titles, recently published, both original Dutch and translated fiction, covering many genres. This includes new editions of older titles. Metadata includes a.o. author, publication year, publisher and genre.
Review corpus	Currently 472,810 Dutch online book reviews, including metadata about the reviewer. We will be adding reviews throughout the project.
Book lists	At present 37,453 book lists, created by individual readers either implicitly (by reviewing) or explicitly, often based on genre or appreciation.

Table 1. The book and review datasets.

Reviewer metadata (1)
Based on reviewers' reviews and book lists we cluster readers to identify different groups in terms of reading preferences, how often they read and review books, whether they read books from many genres or a few, and whether they focus on mostly popular or less known books or a mixture of both.

Review analysis (2)
The hundreds of thousands of reviews provide us with human judgements on thousands of books. However, their format is unstructured text. The challenge is to extract judgements on various literary aspects. To measure the reading impact of a novel, we use our Reading Impact model (Boot & Koolen 2020, Koolen, Boot & van Zundert 2020), which identifies and categorizes sentences expressing impact of narrative or stylistic aspects. Review metadata include the reviewer's rating and reader group. Reader groups can be used to differentiate between types of reader responses. E.g. suspense genre readers may find the story characters of a book boring, while readers of other genres find them intriguing. Where necessary we enrich the reviews with human judgements of specific literary aspects by manually annotating a set of reviews, jump-started by extracting passages from reviews using lists of signal words. Different reviewers of the same book will have different opinions, and so the collection of a book's reviews provides us with a distribution of aspect annotations.

Book analysis (3)
For book analysis we rely on book metadata (genre, publisher, sales and lending figures) and the book text, in which we distinguish between lower-level and higher-level features. Lower-level features are the result of automated analysis tools at the language level, including the output of NLP parsers (e.g. Alpino[1]). Lower-level features include anything from sentence length to a parse tree. From a literary perspective, these are hard to interpret.

Higher-level features are features that can be provided with a literary interpretation on the basis of theories and models such as those relating to narrative arcs, absorption or affective responses. High-level features are built from the output of existing (Dutch)

[1] http://www.let.rug.nl/vannoord/alp/Alpino/

language resources such as lexicons for emotions, personality traits, concreteness and various other psychological and lexical categories via LIWC and T-scan (see Table 2). If the Dutch language resources are not sufficient, we fallback to machine translating the texts to English and using English resources (Ye & Boot 2020). High-level features can also be constructed on top of other features by statistical analysis: the creation of condensed representations that summarize a series of scores of a lower level. Simple examples would be to compute the variance of a feature such as sentence length (Mohseni 2020), or to average emotion scores for the whole book, or to derive a development-profile (similar to Jockers 2015). A book's sadness might for instance follow a low-high-low pattern. We can also combine scores of individual aspects. E.g. does the main character show mostly a single emotion or a more balanced set of emotions?

Lexicon	Description	Reference
Moors-Lexicon	4300 Dutch word-associations for valence, arousal and dominance	Moors et al. 2013
Brysbaert-Lexicon	concreteness and age-of-acquisition norms for 30000 Dutch words	Brysbaert et al. 2014
NRC EmoLex (NL)	9922 Dutch word-associations for 8 basic emotions	Mohammad & Turney 2010, Van Lange & Futselaar 2020
Idioticon van de persoonlijkheid	1186 phrases for five personality traits	De Raad & Doddema-Winsemius 2006
T-Scan / RBN	Dutch word lists for concreteness/abstractness, action verbs, ...	Pander Maat et al. 2014
Linguistic Inquiry and Word Count	Dutch word lists for sentiment, cognitive processes, subjectivity, ...	Pennebaker et al. 2007, Boot, Zijlstra en Geenen 2017
'Personagebank', (Character database)	Crowd-sourced resource for characters in Dutch novels, http://personagebank.nl/	Smeets et al. 2019

Table 2. Lexicons for features relating to emotion, personality, concreteness, action, etc.

Matching book features and reader opinions (4)

As mentioned above, the annotation and analysis at the review level gives us collections of labels at the book level that express the readers' appreciation of the book. The next step is to try to predict the readers' appreciation on the basis of the book features. We will use both ML and statistical inference approaches. We distinguish between the features derived from the impact model and the other review-derived features. We will use the latter to select the features to predict the former. We create ensembles of successful measures e.g. by filtering out highly correlated measures, by brute-force testing, by choosing comprehensible measures over black-box ones.

Example features

To clarify this process, we describe it in more detail for three aspects of fiction, one from each of our three categories (Overall, Stylistic and Narrative). For the other aspects, table 3 provides some pointers to our approach.

Mood

We look at the emotions that literature evokes (Mar et al. 2011) and basic emotional arcs in literary texts (Reagan et al. 2016). Our approach to gauge mood development in literary narrative is to build on top of existing Sentiment Analysis tools such as LIWC. We also use the recent SentiArt approach (Jacobs, 2019). SentiArt uses pre-trained word embeddings to calculate the similarity of any word in a literary text corpus to a list of main emotion words. It builds a 2D emotion space that helps us to produce plots that chart the mood development. We look at correlations between the book's higher-level mood features (e.g. mood development shapes) and the moods annotated in the reviews. We select the best mood features for predicting the impact factors.

Complexity

To create annotations on a book’s complexity we use a combination of book metadata (genre, publisher), attractiveness to reader groups, ratings and commercial success (i.e. no manual annotation). The textual book features that we use include word frequencies, sentence length, POS labels, parse trees, mood and topic. Technologies that we will use for transformations include statistical variance (Mohseni et al 2020), predictability of next sentence and next word using BERTje (a Deep Learning model for Dutch language understanding, De Vries et al 2019) and/or a Recurrent Neural Network (Manjavacas et al 2017), entropy (Febres et al 2017), Zipf distributions (Febres et al 2017), fractality (Mohseni et al 2020) and text-as-network measures (Amancio 2015). Then we match book features and reader opinions.

Attractiveness of characters

Maslej et al. (2019) found that affective themes, abstractness and emotion words scoring high on valence and arousal influence a character’s appeal to readers. To get reader opinions on story characters from the reviews, we will use lists of signal words for characters (main character, protagonist, …) and frequently occurring names (possibly bootstrapped from *Personagebank*, see Table 2) to identify passages in reviews discussing the main story characters and annotate them with attractiveness labels, building on the model of Kim & Klinger (2018). From the book texts, we follow the approaches of Elsner (2012), Nalisnick and Baird (2013), Jacobs (2019) and Cheng (2020) and extract passages in which story characters are mentioned (where first-person narratives create an additional challenge of identifying the narrating character(s)) and score them on arousal, valence, dominance, basic emotions and personality traits using existing lexicons (see Table 2). Using the distribution of the scores across the book, we create high-level features by computing profiles from the individual emotional and personality dimensions.

Feature group	Textual features
Overall	**Topics:** USAS (Rayson et al. 2004), LIWC (Pennebaker et al 2007), topic modelling induced topics.
Stylistic	**Concreteness**: T-scan, Brysbaert-lexicon (see table 2) **Densities**: (of events) T-scan computes the density of verbs related to human acts; (of dialogue) one fruitful approach based on speech act identification (a.o. Qadir & Riloff 2011, Morales-Ramirez & Perini 2014); (of reflection) T-scan computes several categories of verbs and nouns that may express reflection; (of description) measures in T-scan **Unexpectedness and Foregrounding**: deviation w.r.t. 'normal' language use within/across genres, using entropy-based measures, T-scan, fluctuation analysis (Mohseni 2020).
Narrative	**Narrative pace**: Underwood 2018.

Table 3. Technology pointers for the features not explicitly discussed.

Reuse, Sustainability, Dissemination and Collaboration

Re-use: Our results, tools and resources will be of great value for the many Dutch and English speaking researchers cited in the bibliography. For English translations of resources and tools, the potential impact is even bigger. (Several international annual workshops are devoted to this field). Furthermore, we expect our approach will influence theoretical developments in literary research and increase the acceptance of computational approaches as a part of literary research methodology. Our dataset of annotated reviews can be used for ML research on extracting reader judgements.

We expect our approach to be useful to other fields where predicting responses to stories is important, e.g. in film and television studies, communication studies, in research towards response to art more generally, and in (automatic) story-generation research.

Predicting an audience’s response based on measures of (textual) content is an important objective for businesses ranging from publishers, journalism and game developers to creative writing. We work with seven Dutch book publishers interested in software that analyses novels and their impact on readers. Our research is also a step towards reading recommendation systems with obvious interest for them and the National Library.

Sustainability: We will stimulate reuse by providing all code bases and software libraries in open source repositories and resources as open access datasets in data repositories (e.g. Zenodo and DANS). This includes generated metadata and derived data, measurements, and inferred high level features and metrics. In case we apply any closed source solutions in the project, we will include documentation on how to integrate them.

During the project we have support from both the KNAW Humanities Cluster and KNAW IT infrastructure for CPU intensive or cluster computing when needed. The Humanities Cluster is one of the core organizations involved in the Dutch research infrastructure CLARIAH+, through which it can guarantee the long-term maintenance and accessibility of resources.

Dissemination:
We will disseminate the project and its results to researchers through publications, a conference, a research competition, Twitter, a weblog and the various repositories mentioned above. As for the general public, there exists a wide interest in computational research into literature. We expect a continued interest in the research at NRC Handelsblad and other media. We will target other broad audience channels with regular press briefs and inform interested publishers with a six-monthly newsletter.

Collaboration: See under Reuse, last paragraph.

Workplan

We distinguish five phases. After a preparatory phase, three (independent) phases study one feature type, ordered by increasing complexity. Results for each phase are disseminated through one conference paper and one journal article produced during the following phase. The final phase aggregates results and involves the broader research community. We will port a part of the tools to English.
Applicants, post-doc and RSE are jointly responsible for the scientific quality of the project's work. Postdoc and applicants are responsible for embedding the work in current literary scholarship. The postdoc oversees and contributes significantly to the annotation activities, for which we will also hire annotators. The RSE works importantly on extracting low-level features, organising them into higher-level features and testing their suitability in predicting or learning the labels resulting from annotation and the impact factors. He/she also participates in annotation guideline development and contributes a small amount of time towards annotation to develop a shared understanding of all scientific results. The applicants are involved in research design, guidelines development, annotation, writing, preparing the conference and competition, and in writing a follow-up application and evaluation.

Phase	Months	Description
1. startup	1-6	• Compute output variables (Boot & Koolen, in press) • Theorize features (see table 1 in 6b) • Reader clustering • Develop annotation guidelines • Corpus set-up (based on existing collections), prepare pipelines RSE effort: 0.5 FTE
2. Common	7-14	• Select and prepare reviews for annotation • perform topic modelling, sentiment analysis • select or design higher level features based on mood and topic • predict or learn labels resulting from annotation • predict impact factors • Select and annotate reviews for mood, topic, stylistic features RSE effort: 1.0 FTE
3. Stylistic	15-22	• Develop large collection of measures for stylistic features • select or design higher level features for style • predict or learn labels resulting from annotation as well as metadata-based labels • predict impact factors • Evaluate and fine-tune common feature extractors • Annotate narrative features RSE effort: 1.0 FTE
4. Narrative	23-30	• Start translating • Prepare conference and competition • Develop feature extractors and measures for narrative features: recognize characters, collect low-level features,

		deduce higher-level features • predict or learn labels resulting from annotation • predict impact factors RSE effort: 1.0 FTE
5. Wrap-up	31-36	• Based on results of previous phases, develop and test final models for predicting impact • Validation: assessing accuracy, suitability, limitations. • Organising a conference, research competition, proceedings. • Develop English software. • Documenting and publishing code and datasets RSE effort: 0.5 FTE

Deliverables

- A software package that computes relevant features of Dutch fiction and a select number of features for fiction in English, released under MIT license as to enable commercial reuse in possibly derived tools. We implement a proof-of-concept web service within the CLARIAH+ infrastructure.
- 4 scholarly articles (one for each WP starting with WP2) in Poetics, Digital Scholarship in the Humanities, Scientific Study of Literature, Language and Literature
- 4 presentations at these conferences: Digital Humanities, the conference of PALA (Poetics And Linguistics Association), IGEL (International Society for the Empirical Study of Literature), Achter de Verhalen ('Behind the Narrative', series of Dutch-language conferences on modern literature), ...
- A (self-financing) international conference on predicting response to fiction including a competition around a shared task, using the annotations and a selection of texts as the common dataset and evaluation (including a set for which we have both English and Dutch versions of the same novel, aligned at sentence level, for which we will use the annotations created in the project to build multilingual test collections, to evaluate the competing algorithms).
- A bimonthly blog discussing problems, procedures, results, targeted at both researchers, stakeholders, and a broader audience
- Six-monthly newsletter to publishers
- Participation in the DRONGO language festival, to share our findings with an interested general audience (https://www.drongotalenfestival.nl/engels/)
- Regular press briefs that link to our blog and other publications
- A new European grant application seeking to deepen our findings with our European network

References

Amancio, D. R. (2015). A complex network approach to stylometry. PloS one, 10(8), e0136076.

Andringa, E. (1996). Effects of ‘narrative distance’ on readers' emotional involvement and response. Poetics, 23(6), 431-452.

Archer, J., & Jockers, M. L. (2016). The Bestseller Code: Anatomy of the Blockbuster Novel. New York: St. Martin's Press.

Ashok, V. G., Feng, S., & Choi, Y. (2013). Success with style: Using writing style to predict the success of novels. Poetry, 580(9), 70.

Blackburn, K. G. (2015). The narrative arc: exploring the linguistic structure of narrative. (PhD), University of Texas at Austin, Austin.

Bogers, T., Gäde, M., Koolen, M., Petras, V., Skov, M. "What was this Movie About this Chick?" - A Comparative Study of Relevance Aspects in Book and Movie Discovery. In Transforming Digital Worlds - 13th International Conference, iConference 2018, Sheffield, UK, March 25-28, 2018, Proceedings, pages 323--334, 2018.

Bogers, T., & Koolen, M. (2017, August). Defining and supporting narrative-driven recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (pp. 238-242).

Boot, P. (2019). Editing social media: the case of online book discussion. International Journal of Digital Humanities, 1, 251–262.

Boot, P. (2020). The Voice of the Reader: The Landscape of Online Book Discussion in the Netherlands, 1997-2016. In J. Rose (Ed.), The Edinburgh History of Reading. Common Readers. Edinburgh: Edinburgh University Press.

Boot, P. & Koolen, M. (2020). Captivating, splendid or instructive? Assessing the impact of reading in online book reviews. Scientific Study of Literature 10:1, pp. 66–93. issn 2210-4372 | e-issn 2210-4380 DOI: 10.1075/ssol.20003.boo. Pre-pub available at: ssol.20003.boo-proofs-after-correction.pdf

Bortolussi, M., & Dixon, P. (2003). Psychonarratology. Foundations for the Empirical Study of Literary Response. Cambridge: Cambridge University Press.

Bouma, G., Van Noord, G., & Malouf, R. (2001). Alpino: Wide-coverage computational analysis of Dutch. In Computational linguistics in the Netherlands 2000 (pp. 45-59). Brill Rodopi.

Bourdieu, P. (1985). The market of symbolic goods. Poetics, 14(1-2), 13-44.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. Acta psychologica, 150, 80-84.

Burke, M., & Evers, K. (2017). Formalist stylistics. In M. Burke (Ed.), The Routledge handbook of stylistics (pp. 31-44): Routledge.

Burroway, J. (2010). Imaginative writing: The elements of craft: Longman Pub Group.

Cheng, J. Y. (2020). Aspects of Character: Quantitative Evidence and Fictional People. (PhD), The University of Nebraska-Lincoln, Lincoln, Nebraska.

Chin-A-Fo, H & Jaeger, T (2020). Mannen en geweld zijn dé ingrediënten voor een bestseller, zeggen algoritmes. NRC 5 maart 2020.

Crosbie, T., French, T., & Conrad, M. (2013). Towards a model for replicating aesthetic literary appreciation. Paper presented at the Proceedings of the Fifth Workshop on Semantic Web Information Management.

De Nooy, W. (2002). The dynamics of artistic prestige. Poetics, 30(3), 147-167.

De Vries, W., van Cranenburgh, A., Bisazza, A., Caselli, T., van Noord, G., & Nissim, M. (2019). Bertje: A dutch bert model. arXiv preprint arXiv:1912.09582.

Driscoll, B., & Rehberg Sedo, D. (2019). Faraway, So Close: Seeing the Intimacy in Goodreads Reviews. Qualitative Inquiry, 1077800418801375.

Elsner, M. (2012, April). Character-based kernels for novelistic plot structure. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 634-644).

Febres, G., & Jaffé, K. (2017). Quantifying structure differences in literature using symbolic diversity and entropy criteria. Journal of Quantitative Linguistics, 24(1), 16-53.

Gavaler, C., & Johnson, D. (2019). The literary genre effect: A one-word science fiction (vs. realism) manipulation reveals intrinsic text properties outweigh extrinsic expectations of literary quality. Scientific Study of Literature, 9(1), 34-52.

Gutjahr, P. C. (2002). No Longer Left Behind: Amazon.com, Reader-Response, and the Changing Fortunes of the Christian Novel in America. Book History, 5, 209-236.

Harris, S. K. (2019). Whohoo!!! Joan of Arc!!!!! American Literary Realism, 51(2), 136-153.

He, H., Barbosa, D., & Kondrak, G. (2013, August). Identification of speakers in novels. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1312-1320).

Hovinga, E. (2019). De impact van het boek. Een analyse van de impact van lezen op de mens en de maatschappij. Retrieved from https://www.kvbboekwerk.nl/monitor/impact-van-het-boek

Hutto, C.J., Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Jacobs, A. M. (2019). Sentiment analysis for words and fiction characters from the perspective of computational (Neuro-) poetics. Frontiers in Robotics and AI, 6, 53.

Jockers, M. L. (2015). Syuzhet: Extract sentiment and plot arcs from text. Retrieved October, 21, 2015.

Kar, S. (2020). Automatic Characterization of Stories. (PhD), University of Houston, Houston.

Karsdorp, F., Manjavacas, E., Burtenshaw, B., & Kestemont, M. (2017). Synthetische literatuur: neurale netwerken als coauteur. DIXIT: tijdschrift over taal-en spraaktechnologie, 14.

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. Science, 1239918.

Kim, E., & Klinger, R. (2018, August). Who feels what and why? annotation of a literature corpus with semantic roles of emotions. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 1345-1359).

Koolen, M., Bogers, T., Gäde, M., Hall, M., Hendrickx, I., Huurdeman, H., Kamps, J., Skov, M., Verberne, S., Walsh, D. Overview of the CLEF 2016 Social Book Search Lab. In Experimental IR Meets Multilinguality, Multimodality, and Interaction - 7th International Conference of the CLEF Association, CLEF 2016, Evora, Portugal, September 5-8, 2016, Proceedings, pages 351--370, 2016.

Koolen, C., van Dalen-Oskam, K., van Cranenburgh, A., & Nagelhout, E. (2020). Literary quality in the eye of the Dutch reader: The National Reader Survey. *Poetics*, 101439.

Koopman, E. M. E., & Hakemulder, F. (2015). Effects of literature on empathy and self-reflection: A theoretical-empirical framework. Journal of Literary Theory, 9(1), 79-111.

Lendvai, P., Darányi, S., Geng, C., Kuijpers, M., de Lacalle, O. L., Mensonides, J.-C., . . . Reichel, U. (2020) Detection of Reading Absorption in User-Generated Book Reviews: Resources Creation and Evaluation. Paper presented at the The 12th Language Resources and Evaluation Conference, Marseiile, France.

Louwerse, M., Benesh, N., & Zhang, B. (2008). Computationally discriminating literary from non-literary texts. Directions in Empirical Literary Studies: In Honor of Willie Van Peer, 175.

McGillivray, Barbara, Thierry Poibeau, and Pablo Ruiz Fabo. (2020). Digital Humanities and Natural Language Processing: 'Je t'aime... Moi Non Plus.' DHQ: Digital Humanities Quarterly 14(2):454. http://www.digitalhumanities.org/dhq/vol/14/2/000454/000454.html.

Manjavacas, E., De Gussem, J., Daelemans, W., & Kestemont, M. (2017). Assessing the stylistic properties of neurally generated text in authorship attribution. *arXiv preprint arXiv:1708.05536*.

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. Perspectives on psychological science, 3(3), 173-192.

Maslej, M. M., Mar, R. A., & Kuperman, V. (2019). The textual features of fiction that appeal to readers: Emotion and abstractness. Psychology of Aesthetics, Creativity, and the Arts.

Miall, D. S. (2004). Episode structures in literary narratives. Journal of literary semantics, 33(2), 111-129.

Miall, D. S. (2006). Literary reading: empirical & theoretical studies. New York: Peter Lang Publishing.

Miall, D. S., & Kuiken, D. (1999). What is literariness? Three components of literary reading. Discourse Processes, 28(2), 121-138.

Mohseni, M., Gast, V., & Redies, C. (2020). Comparative Computational Analysis of Global Structure in Canonical, Non-Canonical and Non-Literary Texts. arXiv preprint arXiv:2008.10906.

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A.-L., Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. Behavior research methods, 45(1), 169-177.

Morales-Ramirez, I., & Perini, A. (2014, June). Discovering Speech Acts in Online Discussions: A Tool-supported method. In CAiSE (Forum/Doctoral Consortium) (pp. 137-144).

Nalisnick, E. T., & Baird, H. S. (2013, August). Character-to-character sentiment analysis in Shakespeare's plays. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 479-483).

Pander Maat, H., Kraf, R., van den Bosch, A. P. J., Dekker, N., Gompel, M. v., Kleijn, S., . . . Sloot, K. v. d. (2014). T-Scan: a new tool for analyzing Dutch text. Computational Linguistics in the Netherlands Journal, 4.

Pennebaker, J. W., Chung, C. K., Ireland, M. E., Gonzales, A., & Booth, R. J. (2007). The development and psychometric properties of LIWC2007.   Retrieved from http://hdl.handle.net/2152/31333

Qadir, A., & Riloff, E. (2011, July). Classifying sentences as speech acts in message board posts. In Proceedings of thfrae Conference on Empirical Methods in Natural Language Processing (pp. 748-758). Association for Computational Linguistics.

Rayson, P., Archer, D., Piao, S., & McEnery, A. M. (2004). The UCREL semantic analysis system.

Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. EPJ Data Science, 5(article no. 31).

Rebora, S., Boot, P., Pianzola, F., Gasser, B., Herrmann, J. B., Kraxenberger, M., . . . Messerli, T. C. (2019). Digital Humanities and Digital Social Reading. OSF Preprint, https://osf.io/mf4nj

Reuver, M. (2019). Amputation or Accident? Classifying Dutch Urban Legends into Story Types with a Hierarchical Classifier. DH Benelux 2019, Liége, Belgium. http://2019.dhbenelux.org/wp-content/uploads/sites/13/2019/08/DH_Benelux_2019_paper_69.pdf

Riddell, A., & van Dalen-Oskam, K. (2018). Readers and their roles: Evidence from readers of contemporary fiction in the Netherlands. PloS one, 13(7).

Rosenblatt, L. M. (1994). The reader, the text, the poem: The transactional theory of the literary work: Southern Illinois Univ Pr.

Saricks, J. G. (2005). Articulating a Book's Appeal. In: Readers' advisory service in the public library. Third Edition (pp. 40-73). Chicago: American Library Association.

Smeets, R., Sanders, E., & van den Bosch, A. (2019). Character Centrality in Present-Day Dutch Literary Fiction. Digital Humanities Benelux Journal, 1(1).

Towheed, S., Crone, R., & Halsey, K. (Eds.). (2011). The history of reading : a reader: Routledge.

Tulkens, S., van Cranenburgh, A. (2020). An unsupervised aspect extraction method with an application to Dutch book reviews. Computational Linguistics in the Netherlands 2020 (CLIN 2020).

Underwood, T. (2018). Why Literary Time Is Measured in Minutes. ELH, 85(2), 341-365.

Van Cranenburgh, A., & Bod, R. (2017). A data-oriented model of literary language. arXiv preprint arXiv:1701.03329.

Van Cranenburgh, A. (2019). A Dutch coreference resolution system with an evaluation on literary fiction. *Computational Linguistics in the Netherlands Journal*, *9*, 27-54.

Van Cranenburgh, A., van Dalen-Oskam, K., & van Zundert, J. (2019). Vector space explorations of literary language. Language Resources and Evaluation, 53(4), 625-650

Van den Bosch, A., Busser, G.J., Daelemans, W., and Canisius, S. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch, In F. van Eynde, P. Dirix, I. Schuurman, and V. Vandeghinste (Eds.), Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting, Leuven, Belgium, pp. 99-114

Van den Hoven, E., Hartung, F., Burke, M., & Willems, R. (2016). Individual differences in sensitivity to style during literary reading: Insights from eye-tracking. Collabra: Psychology, 2(1).

Van der Deijl, L., Pieterse, S., Prinse, M., & Smeets, R. (2016). Mapping the Demographic Landscape of Characters in Recent Dutch Prose: A Quantitative Approach to Literary Representation. Journal of Dutch Literature, 7(1).

http://www.journalofdutchliterature.org/index.php/jdl/article/view/103.

Van Zundert, J., Koolen, M., & van Dalen-Oskam, K. (2018). Predicting Prose that Sells: Analyzing Bestseller Prediction Models. Paper presented at Data in Digital Humanities. European Association for Digital Humanities 2018.

Visser, N., Koolen, C. (2019). Identifying the relation between author gender and Dutch suspense novels. DH Benelux 2019, Liége, Belgium. http://2019.dhbenelux.org/wp-content/uploads/sites/13/2019/08/DH_Benelux_2019_paper_41.pdf

Wilhelm, J. D. (2016). "You gotta be the book": teaching engaged and reflective reading with adolescents: Teachers College Press.

Wimsatt, W. K. (1954). The Affective Fallacy. In: The verbal icon: Studies in the meaning of poetry (pp. 21-39): University Press of Kentucky.