

## Thinking Outside of the Bounding Box: A Reconsideration of the Application of Computational Tools on Uncertain Humanities Data

Recent debate about what to do with books that present outdated or disfavored views has shown that literary representation is a highly controversial issue. Children's books often depict a fictional world, so one could wonder why the content of those works matters at all. The answer lies in the fact that books for children reflect specific worldviews. This turns the process of reading into a contact zone, where readers engage with different ideas about themselves and the world (Čermáková & Mahlberg, 2021). It is a dynamic process, where some social norms are internalized, while others are not (Tatar, 2009; Brown, 2017). The books thus influence to a significant extent how (young) readers perceive reality, themselves, and others (Van den Bossche & Klomberg, 2020). Despite the significance of representation in books for children, we are still missing key information on how it is given shape exactly.

This project works towards a better understanding of the historical evolution of representation along the lines of age, race, class, and gender. It does so by analyzing a diverse corpus of 1,000 illustrated children's books published in Dutch from the period 1800-1940. Taking a digital approach to this question helps us to overcome the disciplinary canon and to provide a general overview of representation in historical Dutch children's literature, something which is sorely missing. Furthermore, it allows us to analyze how results from a Dutch-speaking context compare to studies based on other language areas. In doing so, this project answers the call by Mansour and Martin for studies situated in different cultures to become conversant with one another (2020). This form of collaborative action can help provide a solid basis for the discussions on representation that are taking place in our current societies.

When studying those materials, however, one is confronted with vagueness on multiple levels. Scholars such as Edmond have already pointed out that this vagueness is one of the defining characteristics of humanities data, which are not subject to verifiable, durable laws (2018). In this presentation I will discuss how working with digital tools – which are often created by computer scientists who had entirely different goals in mind – can be complicated by the nature of humanities' data. The goal is therefore not to present the actual results of this research project, but rather to extend the conversation on how we can bridge the digital and the humanities in DH projects, in a way that is mutually enriching to both fields. The question I will try to answer is: what can we gain from image annotations for object recognition in humanities research, and how does its implementation differ from the more technically oriented use in Machine Learning?

To do so I will focus on uncertainty as a valuable research result, rather than purely a limitation. Following Martín-Rodilla and Gonzalez-Perez I make the distinction between ontological vagueness, or imprecision, and epistemic vagueness, which can be described as uncertainty, and provide details on how we have dealt with both in the implementation of this research (2018). During the first phase of this research, a team of annotators has drawn bounding boxes around all the human characters in the illustrations and assigned them attributes according to their age, race, class, and gender. Through an examination of the inter-annotator agreement, I will surmise what the differences between annotators - due to epistemic vagueness - can entail for our data. In the second part of the presentation, I will

focus on ontological vagueness in the categories we have selected and consider how those choices influence the results of this research.

### References:

- Brown, M. (2017). Children's literature matters. *English Academy Review*, 34(1), 8-22.
- Čermáková, A., & Mahlberg, M. (2021). Gender inequality and female body language in children's literature. *Digital Scholarship in the Humanities*, 36, ii72-ii77.
- Edmond, J. (2018, October). Managing uncertainty in the humanities: Digital and analogue approaches. *Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 840-844.
- Hamilton, M. C., Anderson, D., Broaddus, M., & Young, K. (2006). Gender stereotyping and under-representation of female characters in 200 popular children's picture books: A twenty-first century update. *Sex roles*, 55(11), 757-765.
- Mansour, N., & Martin, M. H. (2020). What Can Danish Multicultural Children's Literature and African American Children's Literature Learn from Each Other? Literary Histories in Dialogue. *Barnboken*, 43.
- Martín-Rodilla, P., & Gonzalez-Perez, C. (2018, October). Representing imprecise and uncertain knowledge in digital humanities: a theoretical framework and ConML implementation with a real case study. *Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 863-871.
- Van den Bossche, S., & Klomberg, A. (2020). *Jeugdliteratuur door de lens van etnisch-culturele diversiteit*. Eburon.
- Tatar, M. (2009). *Enchanted hunters: The power of stories in childhood*. WW Norton & Company.