Style characterization of machine printed texts
Bagdanov, A.D.

Citation for published version (APA):
Bagdanov, A. D. (2004). Style characterization of machine printed texts

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
They all fall there so perfectly, 
It all seems so well timed. 
And here I sit so patiently
Waiting to find out what price 
You have to pay to get out of
Going through all these things twice.

—Bob Dylan, Stuck Inside of Mobile with the Memphis Blues Again

How are you reading this dissertation?
We tend to take for granted the amount of prior knowledge we apply to the task of decoding the content of a document. Maybe you were curious about the subject of document style characterization. If you are unfamiliar with the subject of document understanding, perhaps you deliberately sought out the introduction, implicitly aware of the existence of chapters, that their beginning is styled differently. How did you find the introduction?

Introductory chapters tend to be toward the beginning. Perhaps you found the table of contents and the exact page number, or maybe you paged through the book searching for pages that look like chapter beginnings. You might have had a mental template of beginning-of-chapter pages similar to the scheme annotating this very page. These types of mental templates are learned through experience with a document style, or genre. The existence of genres of texts establishes implicit rules. Experience with the rules yields knowledge, allowing you to navigate the structure of books, identify the purpose of chunks of text, and find information. How did you anticipate that this would be an italicized interrogative?

Now you are reading this sentence. Focus is subconsciously attenuated to the baseline of these decorated vertical bars of ink. Without having to read anything in this paragraph, your attention is drawn to words that are **boldfaced**, **SMALLCAPPED**, or otherwise **emphasized**. Visual cues attenuate your focus, under the learned prior assumption that emphasized text is important. The author is exploiting this by letting style coincide with focus.

This dissertation examines elements of style in machine printed texts and proposes tools and techniques to characterize them.