

Exploring the similarities and differences between VLM-driven and traditional OCR for Historical Swedish Data



SPRÅKBANKENTEXT

Martin Johansson, Selma Waginder, Dana Dannélls

Department of Computer Science and Engineering,
Department of Swedish, multilingualism, language technology
Chalmers University of Technology, University of Gothenburg

Abstract

Recent Swedish OCR research has focused mainly on traditional approaches, such as CNN–LSTM hybrids and transformer-based models, with limited exploration of vision–language model (VLM) OCR for historical texts. This study investigates how transformers and VLMs compare in character- and word-level recognition across typefaces, and examines their differing error patterns. Fine-tuned Qwen3-VL models (8B and 2B), combined with a simple repetition-trimming step, outperform conventional OCR systems. Remaining errors are largely linked to challenges with Blackletter typefaces and formatting inconsistencies, such as incorrect line breaks, spacing, or character insertions, which significantly impact overall transcription accuracy.

Data



An excerpt from Umebladet, 14th December 1861, with added markings and enumerations to separate segments.

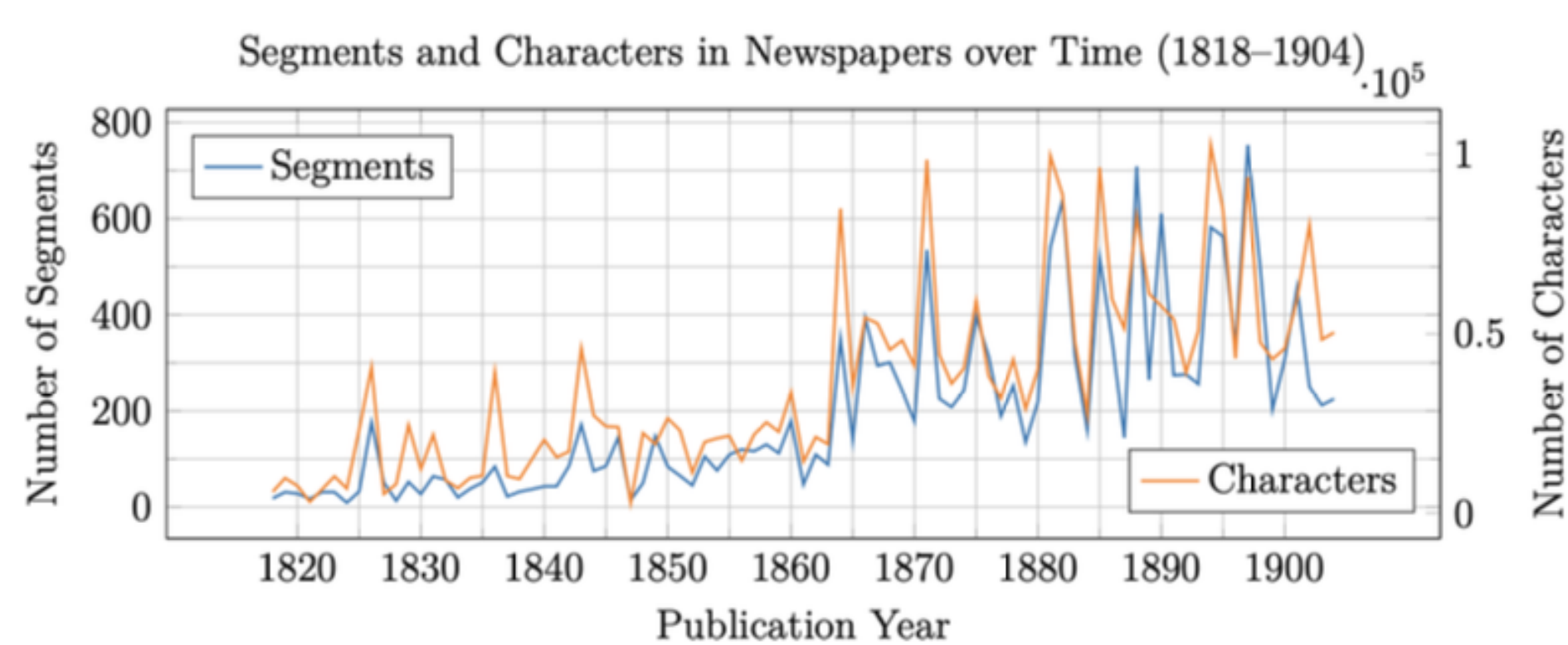
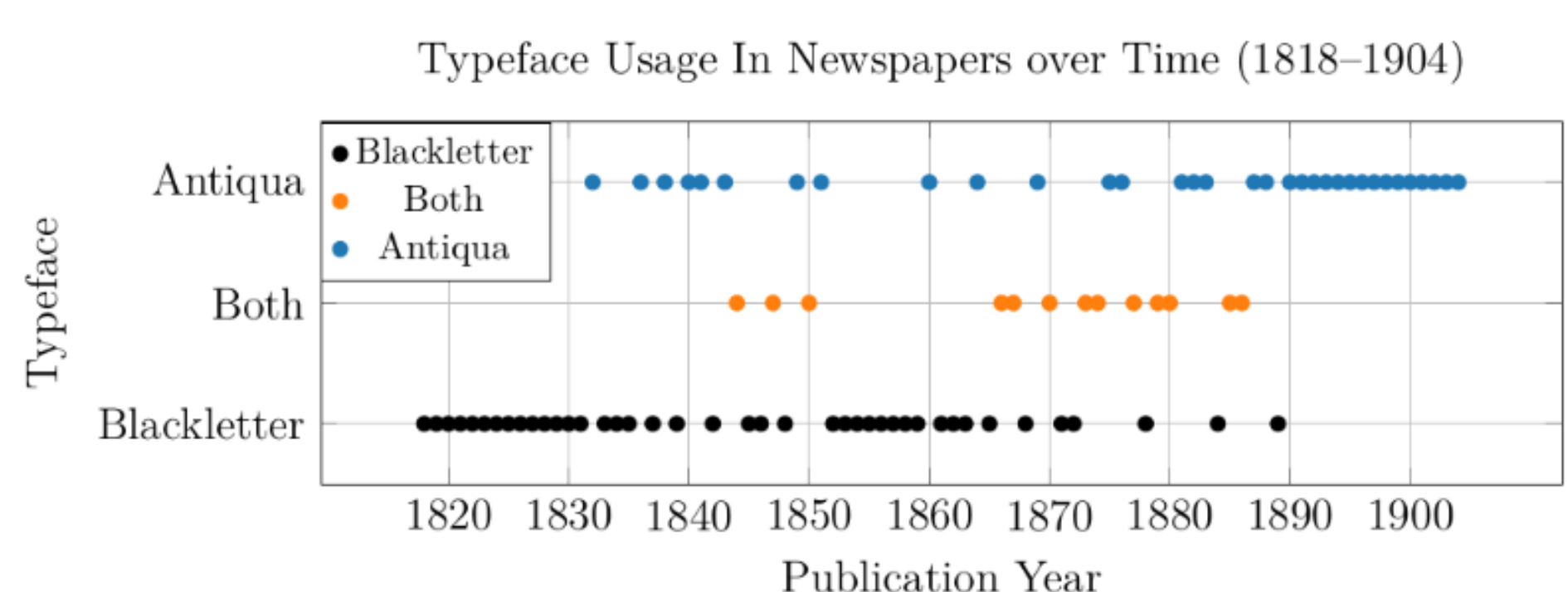
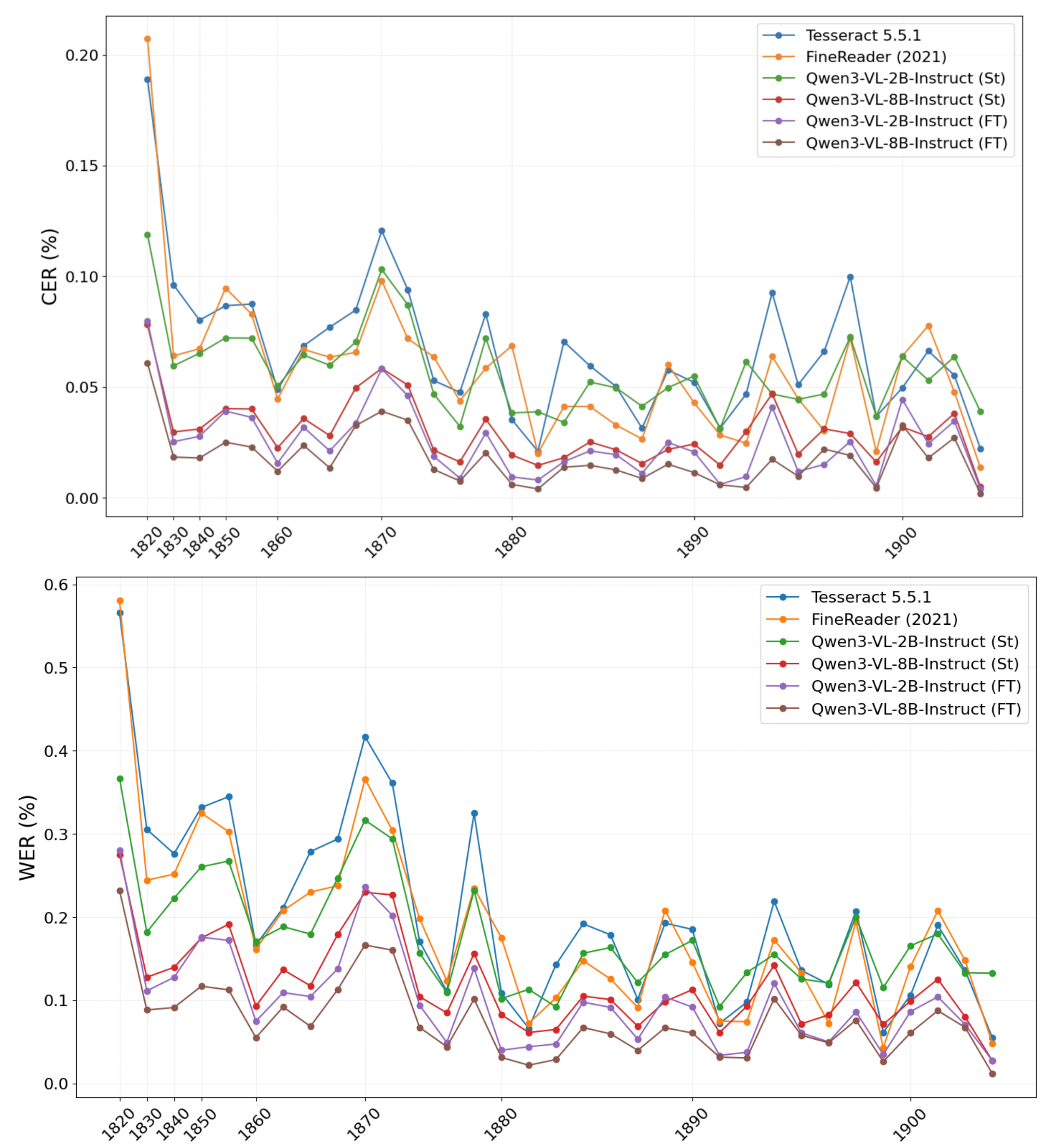


Figure 1: Combined visualization of number of segments and characters in newspaper pages from 1818 to 1904 according to our dataset.



Typeface usage in newspapers over time. Newspapers are labeled “Both” if both typefaces are used in multiple segments.

Experiments and Results



Qualitative Evaluation

Model	Typeface	CER (%)	WER (%)
Qwen3-VL-8B-Instruct (FT)	Blackletter	3.332	14.726
	Antiqua	1.292	5.010
	Both	1.847	8.158
	Total	1.930	8.108
Qwen3-VL-2B-Instruct (FT)	Blackletter	4.646	19.373
	Antiqua	1.802	6.355
	Both	2.673	11.735
	Total	2.707	10.695
Qwen3-VL-8B-Instruct (St)	Blackletter	4.955	20.836
	Antiqua	2.381	8.375
	Both	3.157	13.052
	Total	3.198	12.453
Qwen3-VL-2B-Instruct (St)	Blackletter	8.562	28.820
	Antiqua	4.691	13.145
	Both	6.278	19.210
	Total	5.989	18.304
FineReader (2021)	Blackletter	10.254	35.274
	Antiqua	4.127	12.231
	Both	6.280	21.598
	Total	6.123	19.888
Tesseract 5.5.1	Blackletter	11.242	40.060
	Antiqua	5.033	12.475
	Both	7.417	25.175
	Total	7.089	21.882

Acknowledgments

This work has been funded by Språkbanken – jointly funded by its 10 partner institutions and the Swedish Research Council (2025–2028; project id 2023-00161), and Huminfra, the Swedish Research Council (2021-00176, 2023-00171).