



Style Detection for Free Verse Poetry from Text and Speech

Timo Baumann
Carnegie Mellon University
Language Technologies Institute

Hussein Hussein, Burkhard Meyer-Sickendiek
Freie Universität Berlin
Department of Literary Studies

ABSTRACT:

Modern and post-modern *free verse poems* feature a large and complex variety in their poetic prosodies that falls along a continuum from a **more fluent to a more disfluent and choppy style**.

We investigate the **free verse spectrum** of modern and postmodern poetry, which overcame rhyme and meter, and which instead can be analyzed along our **fluency continuum**.

We present a method, grounded in philological analysis and current research on cognitive (dis)fluency, for automatically analyzing said spectrum.

We define and relate six classes of poetic styles (ranging from *parlando* to *lettristic decomposition*) by their gradual differentiation and present a model for automatic prosodic classification of spoken free verse poetry using both the source text and audio.

The Prosody of Free-verse Poetry

At least 80 per cent of modern and postmodern poems have no rhyme nor metrical schemes such as iambic or trochaic meter. Does this, however, mean that they lack any rhythmical features?

In contrast, the opposite is true: modern poets like Whitman, the Imagists, the Beat poets, and contemporary Slam poets developed a **post-metrical idea of prosody** that employs rhythmical features of everyday language, prose, and musical styles including jazz and hip hop.

An explanation of modern art is the **theory of aesthetic pleasure**, which claims that fluency of cognitive processing causes positive experience. On the other hand, **modern artists often complicate** strategically the processing of their works, possibly to encourage analytical experience and more careful processing on a higher, disfluent, level of abstraction.

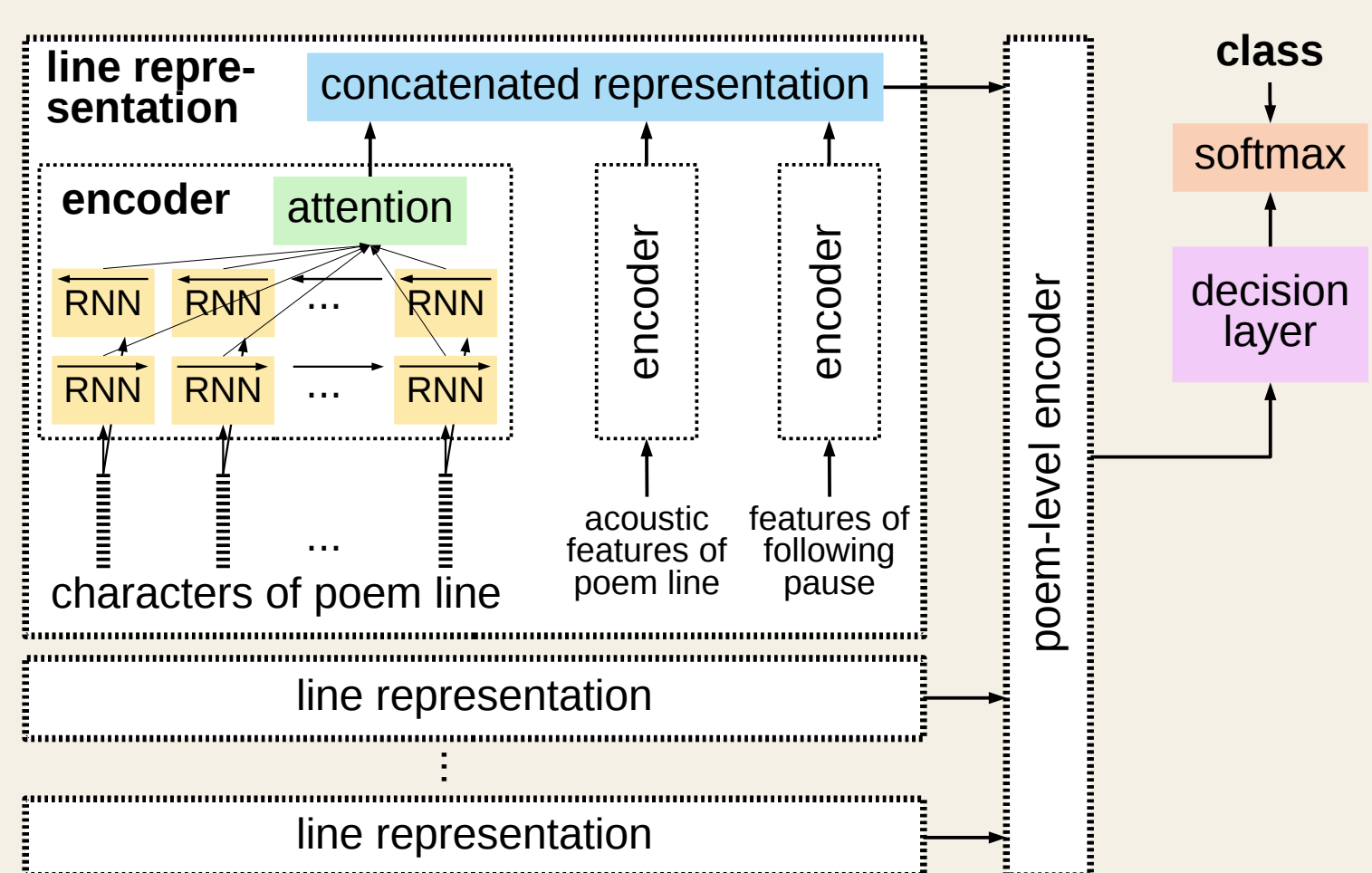
We define the following classes along the fluency spectrum:

- parlando style:** mostly fluent reading of colon-based lines
- variable foot:** emphasis on the gap to the run-on-line
- unemphasized enjambment:** enjambments are spoken but not stressed
- gestic rhythm:** emphasis on 'hard' enjambements
- syllabic decomposition:** dadaistic *sound poetry* with syllables as base
- lettristic decomposition:** sound and visual decomposition of the text



Model

We build a hierarchical neural network (HNN) for style classification:



- main **encoder block:** BiGRU RNN with attention
- encode line **text** based on character encodings
- encode line **acoustics** based on MFCCs & FFVs
- also encode **acoustics of pause** before next line

- concatenate into line representation
- **hierarchically encode** line-by-line representations to **poem representation**
- classification decision based on poem representation
- attention helps **understand model** classification decisions

Ablation: text-only model, model without acoustics of the pause

Analysis of representations: low-dimensional bottleneck (1D/2D/3D) to test hypothesized fluency continuum.

Data Sources/Material

We collaborate with *lyrikline.org*, a website containing hundreds of hours of author-spoken poetry; the German sub-corpus contains 52 hours.

The third author manually assigned a small subset of 175 poems to their rhythmical class as shown below.



	poems	lines	characters	audio
lyrikline: German subcorpus	2392	61849	2025484	52 h
parlando	34	1435	44323	67 min
variable foot	34	878	23684	39 min
unemphasized enjambment	36	1090	33178	48 min
gestic rhythm	33	897	27741	44 min
syllabic decomposition	21	540	12390	26 min
lettristic decomposition	17	684	10007	31 min
deutschestextarchiv.de	—	34291	996714	—

Implementation

text input is via **character embeddings**
forced alignment of speech+text, manual alignment where software fails
extract **MFCCs** and **FFVs**, **z-normalize** each dimension
mean/stddev aggregation every **10 frames**

We implement our neural model in *dyNet*.

- To increase number of training instances, we
- first train a line-by-line encoder and decision layer that classifies every line (5524 instances)
- then train a poem-level encoder and decision layer for all poems (175 instances)

We perform 25-fold cross-validation.

Classification Results

	f-measure	parlando	var. foot	unemph. enj.	gestic	syll. dec.	lettr. dec.
parlando	0.83	30	2	2			
variable foot	0.60	3	20	6	5		
unemph. enj.	0.71	2	4	27	3		
gestic rhythm	0.68		6	5	21		1
syllabic dec.	0.81	2	1			17	1
lettristic dec.	0.77	1				4	12

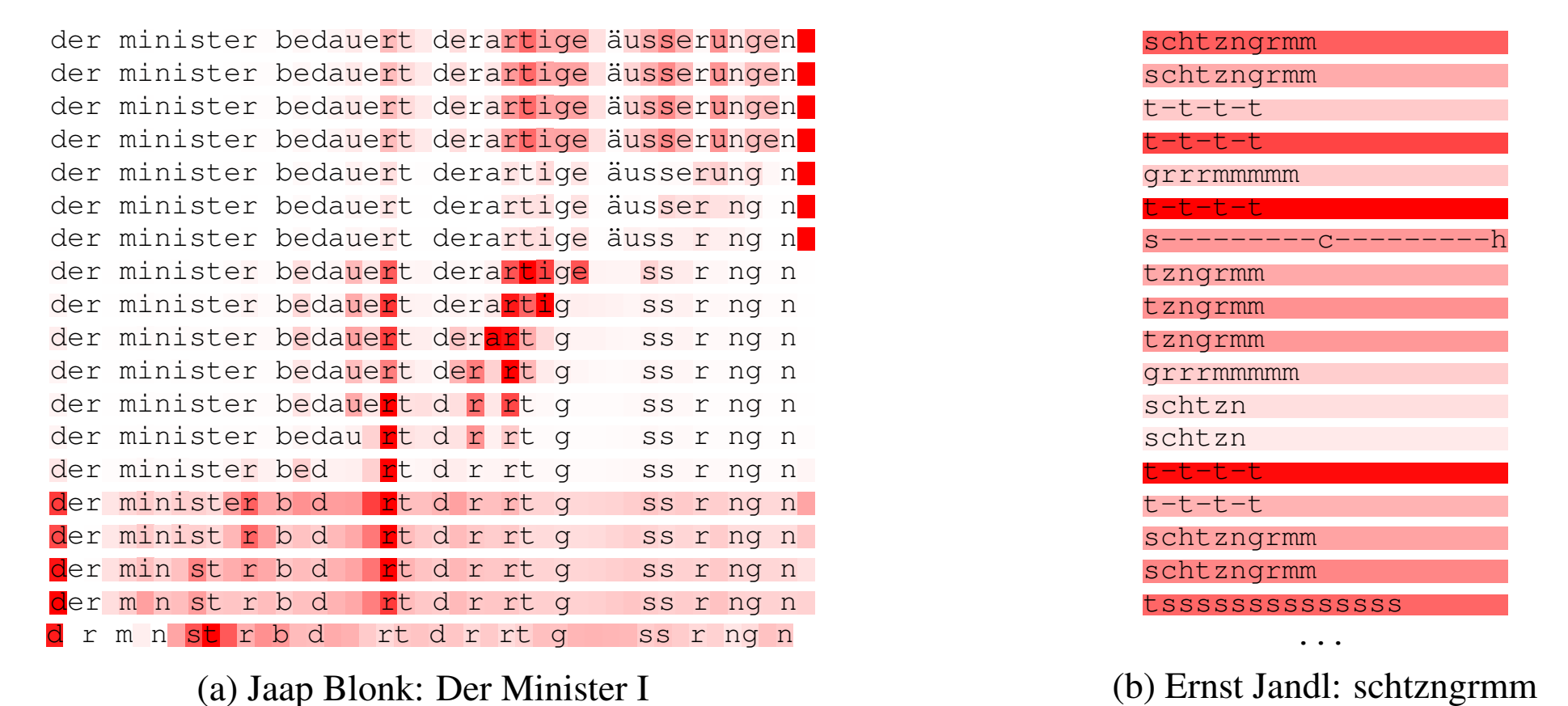
- good performance across most classes (avg. f-measure: .73)
- confusion matrix supports hypothesized fluency continuum

	all features	no pause	text-only
all six classes	0.73	0.66	0.47
parlando vs. variable foot	0.85	0.85	0.65
unemphasized enjambment vs. gestic rhythm	0.78	0.66	0.57 ^{ns}
syllabic vs. lettristic dec.	0.82	0.92	0.82

- ablation shows importance of features in particular for some class differentiations (e.g. pauses help to differentiate enjambment style)

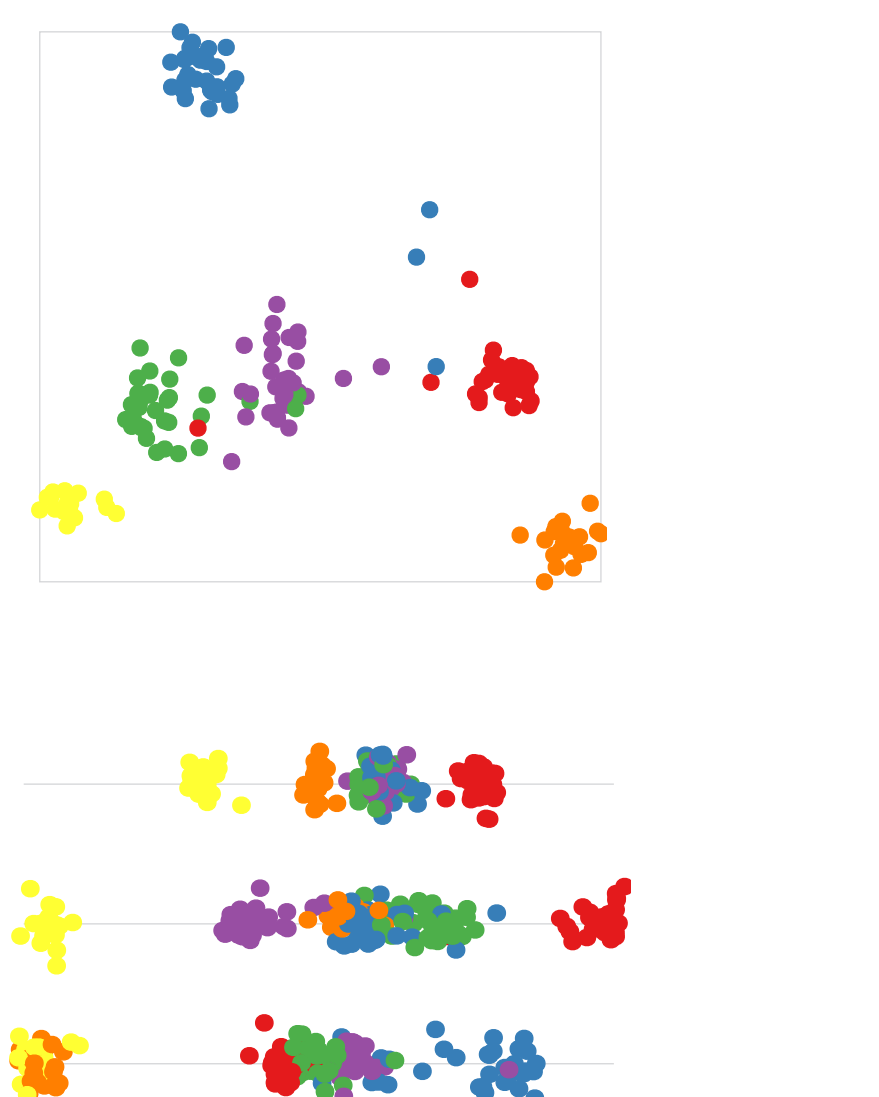
Model Analysis

Attention can partially be interpreted wrt. to decision outcome (e.g. focus of model somewhat on lettristic decomposition boundary, "most" lettristic lines are focused in (b))



Plots of **bottleneck representations**:

- good separation of poetic classes in 2D
- partial confirmation of fluency continuum using 1D representations



Please e-mail us for further information:

- tbaumann@cs.cmu.edu
- hussein@zedat.fu-berlin.de
- bumesi@zedat.fu-berlin.de

<http://www.rhythmicalizer.net>



We gratefully acknowledge funding by a Volkswagen Foundation grant within the *Mixed Methods in the Humanities* program.

