Analysis and Classification of Prosodic Styles in Post-modern Spoken Poetry

Burkhard Meyer-Sickendiek[1], Hussein Hussein[1], Timo Baumann [2]

Department of Literary Studies, Freie Universität Berlin, Berlin, Germany[1]
Language Technologies Institute, Carnegie Mellon University, Pittsburgh, USA[2]

Abstract

We present our research on computer-supported analysis of prosodic styles in post-modern poetry. Our project is unique in making use of both the written as well as the spoken form of the poem as read by the original author. In particular, we use speech and natural language processing technology to align speech and text and to perform textual analyses. We then explore, based on literary theory, the quantitative value of various types of features in differentiating various prosodic classes of post-modern poetry using machine-learning techniques. We contrast this feature-driven approach with a theoretically less informed neural networks-based approach and explore the relative strengths of both models, as well as how to integrate higher-level knowledge into the NN. In this paper, we give an overview of our project, our approach, and particularly focus on the challenges encountered and lessons learned in our interdisciplinary endeavour. The classification results of the rhythmical patterns (six classes) using NN-based approaches are better than by feature-based approaches.

1 Prosodic Styles in Post-modern Spoken Poetry

The most important development in modern and postmodern poetry is the replacement of traditional meter by new rhythmical features. Ever since Walt Whitman's Leaves of Grass (1855), modern (nineteenth- to twenty-first-century) poets have been searching for novel forms of prosody, accent, rhythm, and intonation. Along with the rejection of older metrical units such as the iamb or trochee, a structure of lyrical language was developed that renounced traditional forms like rhyme and meter. Poetic movements and groups like the Imagists (Beyers, 2001; Cooper, 1998; Silkin, 1997), the Black Mountain poets (Berry, 1997; Finch, 2000; Steele, 1990), as well as European poets before and after the second world war (Meyer-Sickendiek, 2012) developed a post-metrical idea of prosody that employs rhythmical features of everyday language, prose, and musical styles including jazz and hip hop. In parts, they intended to create a fluent, in parts a disfluent prosody, for example in Dadaistic Poetry. The text-based patterns include "end-stopped lines," "line-sentences," as developed by Ezra Pound in *Cathay*; dismemberment of the line; spatial dismemberment of the line by indentation, as William Carlos

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2018 in
M. Burghardt C. Müller-Birn (Hrsg.): INF-DH 2018 – Workshopband, 25. Sept. 2018, Berlin
 https://doi.org/10.18420/infdh2018-09

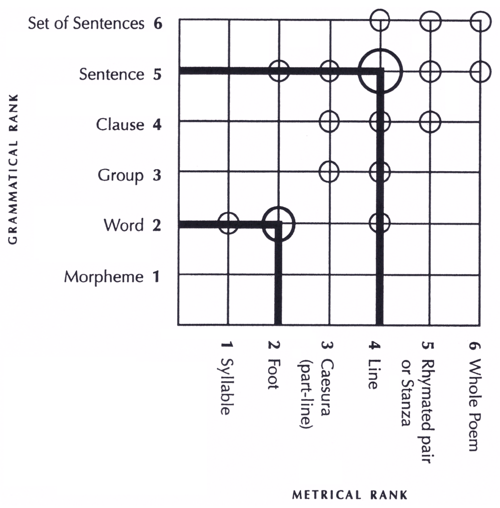

Figure 1: The vertical axis is the grammatical rank; the horizontal axis is the metrical rank. Intersection points help to identify the poem, for instance, the line arrangement (Wesling, 1996).

Williams does in his triadic line verse; systematic enjambment (breaking a sentence or phrase into two lines); or dismemberment with enjambment of the line.

To classify these different features in the line arrangement of each poem, we make use of the theory of the grammetrical ranking. The term grammetrics, coined by Donald Wesling, is a hybridization of grammar and metrics: the key hypothesis is that the interplay of sentence-structure and line-structure can be accounted for more economically by simultaneous than by successive analysis (Wesling, 1996). In poetry as a kind of versified language, the singular sentence interacts with verse periods (syllable, foot, part-line, line, rhymed pair or stanza, whole poem), a process for which Wesling finds ‘scissoring’ an apt metaphor, assuming that meter and grammar can be scissored by each other, that the cutting places can be graphed with some precision (Wesling, 1996, p. 67). In Weslings scheme (see Figure 1), the vertical axis designates the grammatical rank and the horizontal axis the metric rank. Intersections of the two axes are represented by circles in which the axes meet; small circles for small coordinate points, large circles for large ones. Of all possible intersections on the grid, only 16 points are encircled, because normally only these 16 points are filled in poems. The 16 shear points can be represented in two clusters defined by their centers at the main coordinates of word foot and sentence line (large circles).

To detect the specific phrasing of each of these grammetrical units, we make use of Richard Curetons theory of *Rhythmic Phrasing in English Verse* (RPEV) (Cureton, 1992). Cureton proved that the phrasing is a typical feature for poetry-readings. Following Weslings theory of grammetrical ranking, Richard Cureton has divided the poetic rhythm into further components (Cureton, 1992, p. 124) as follows: 1) The *grouping structure* is based on the hierarchical segmentation of a poem, similar to the grammetrical ranking mentioned. 2) The *metrical structure* identifies the regular alternation of strong and weak beats at a number of hierarchical

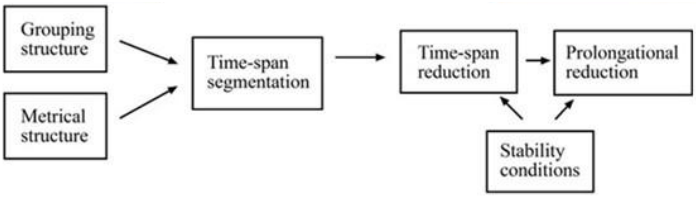

Figure 2: Rhythmic phrasing in English verse (Cureton, 1992).

levels, differing between the beat and the time span between two beats. Both two structures explain the so-called "time-span segmentation". 3) The *time-span reduction* combines the information gleaned from these metrical and grouping structures. This is illustrated in a tree structure-style hierarchical organization uniting time-spans at all temporal levels of a poem. 4) The *prolongational reduction* provides the prosodic phrasing of tensing and relaxing patterns in a readout poem. The prolongation refers to the anticipation and overshooting of a goal, such as the end of a line in an enjambment.

So, all in all our classification pursues two basic questions. (1): Can the poem be determined grammetrically? We use a scheme developed by Donald Wesling (Figure 1), which combines grammatical and metrical ranking. (2): are these grammetric patterns isochronous? In addition to grammetrical ranking, we use the scheme used by Richard Cureton (Figure 2). Based on these two theories, we distinguish 18 rhythmical patterns based on three types of patterns as shown in Figure 3: (a) text-based patterns (unemphasized enjambments, sprungrhythm, parlando, rubato, long line poems, and free associations), (b) audio-based patterns (emphasized enjambments (gestic rhythm), syncopations, variable foot, cut-up rhythm, cadence, and flows), and (c) decompositions (lettristic decompositions, syllabic decompositions, staccato, ellipses, permutations, and dialect poems). We expect that these 18 patterns represent the whole range of free verse prosody in poems read by the authors. For this reason, we establish a gradual one-dimensional continuum, whose two poles are denoted by the terms *fluent* and *disfluent*. The prosodic spectrum with 18 poetic styles in the free verse spectrum is illustrated in Figure 3.

In this paper we analyze the following 6 classes along the fluency/disfluency spectrum: The first two are based on "soft enjambments". The (1) "parlando style" allows a mostly fluent reading of colon-based line-arrangements of the poem. A similar structure is used in the (2) "variable foot", but this pattern emphasizes each gap to the run-on-line. A more disfluent style occurs in the (3) "unemphasized enjambment", when they are strong (not colon-based, but separating articles and nouns or even prefixes of words). When these strong enjambments are emphasized, we got the pattern (4) "gestic rhythm". Two further patterns occur in modern "sound poetry": The (5) "syllabic decomposition", used since dadaism combing syllables; and the (6) "lettristic deiomposition", based on the combination of consonants.

2 Database

The data used in the research project is from the internet portal *lyrikline*. This portal offers over 12,000 audio poems of international poetry, read by the original authors. Users can listen to the poet and read the poems both in their original language and in various translations. The digital material contains more than 10,000 poems by over 1,200 international poets from almost 80 different countries. We examine the German and English poems, with a total of 369 poets (215 German & 154 English) with about 3,600 poems. We expect that most of these poems on *lyrikline* are based on those 18 rhythmical patterns of free verse poetry mentioned above.

3 Modeling Approaches

We build classification of prosodic style based on models of poems. Our goal here is to have a quantifiable measure of success in style modeling, in particular, f-measure of classification performance. We can then measure the influence of various modeling parameters on classification performance. We follow two approaches to model the prosodic styles of poems towards classification, one based on traditional feature engineering in combination with 'readable' classification algorithms such as decision trees. The second approach uses a hierarchical neural network (NN) that encodes the poem into a multi-dimensional representation which is used as input for the final classification. We perform forced-alignment of text and speech for the poems using the text-speech aligner published by (Baumann, Köhn, et al., 2018) which uses a variation of the SailAlign algorithm (Katsamanis et al., 2011) implemented via Sphinx-4 (Walker et al., 2004). The alignments are stored in a format that guarantees the original text to remain unchanged which is important to be able to recreate the exact white-spacing in the poem and would be helpful when adding further annotations (e.g. parts of speech, syntax or semantics) to the poem in the future. We extract the line-by-line timing (start of first and end of last word of the line) for each line.

3.1 Feature-based Classification

On the textual side, we detected the syntactic features, in particular the words' Part-of-Speech (PoS) by using the Stanford parser (Rafferty and Manning, 2008) to parse the written text of poems (parsing each line in isolation). Different features including pause and parser information are used in the classification process (see Hussein et al., 2018a and Hussein et al., 2018b for more details). Three feature sets are utilized: (1) The pause feature set contains two features (the the average pause length at the end of each line as well as between words). (2) Based on the parser output, we compute five features (the poem's number of lines, number of lines with a finite verb, number of lines with punctuation, number of lines with foreign language material, and number of lines with non-words which contain special characters) as the parser feature set. (3) The pause & parser feature set includes seven features which are a combination of the previous pause as well as parser features. We experimented with several classification algorithms (*AdaBoostM1*, *IBk*, *SimpleLogistic*, and *RandomTree*) in the WEKA toolkit (Hall et al., 2009).

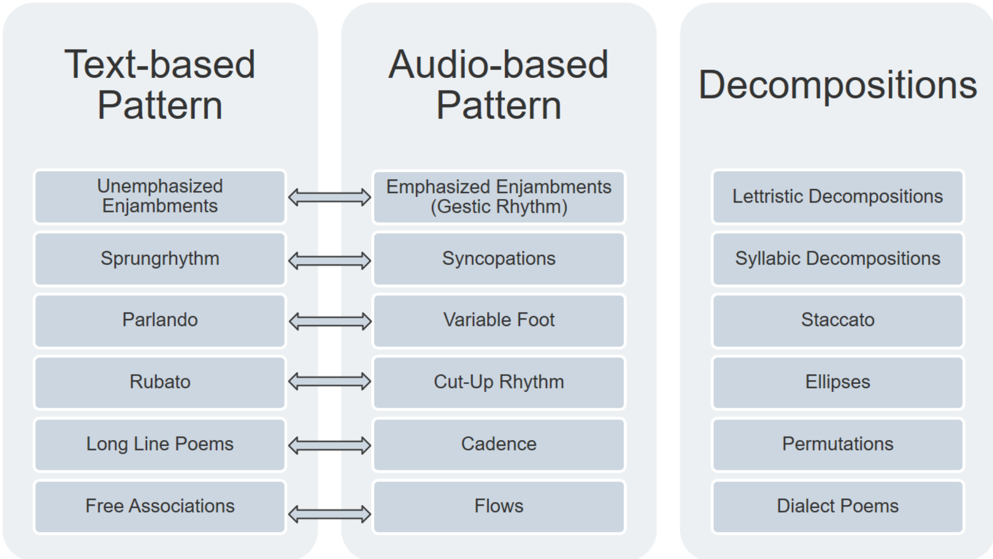

Figure 3: Classification of rhythmical patterns in 3 categories: text-, audio-based patterns, and decompositions.

3.2 Neural Network-based Model

In this section, we describe the performance of the modeling approach based on neural networks for prosodic style classification, with a particular focus on the insights in the humanities that can be derived from the model performance. Given the broad variety of the poems in combination with their relatively small number (see above), our model must deal well with *data sparsity*, i.e. use as few free parameters as possible that need to be optimized during training. For this reason, we decide to focus our textual processing on character-by-character encoding of the lines in the poem (and using character embeddings). We use a bidirectional recurrent neural network (RNN, using gated recurrent unit (GRU) cells (Cho et al., 2014)) which encodes the sequence of characters into a multi-dimensional representation that, although it is not directly applicable to human interpretation, is trained to be optimal towards differentiating the prosodic classes. Our model is not trained using an explicit notion of words. Instead, it may implicitly encode word-level information (such as parts of speech) via the constituting sequences of characters. This is in line with recent work on end-to-end learning, e.g. for speech recognition (Graves and Jaitly, 2014; Hannun et al., 2014), which no more explicitly models phonemes nor words, but directly transfers audio features to character streams. While processing on the word level might allow our model to build a better higher-level understanding of the poem's meaning, this semantic information would likely not help in style differentiation. In addition, word representations would not capture the usage of whitespace, e.g. for indentation, to create justified paragraphs, or other uses, nor special characters. We combine the line-by-line representations using a poem-level encoder which is fed to a decision layer and a final softmax to determine the poem's class, yielding the hierarchical attention network as shown in Figure 4 (see Baumann, Hussein, et al., 2018 for more details).

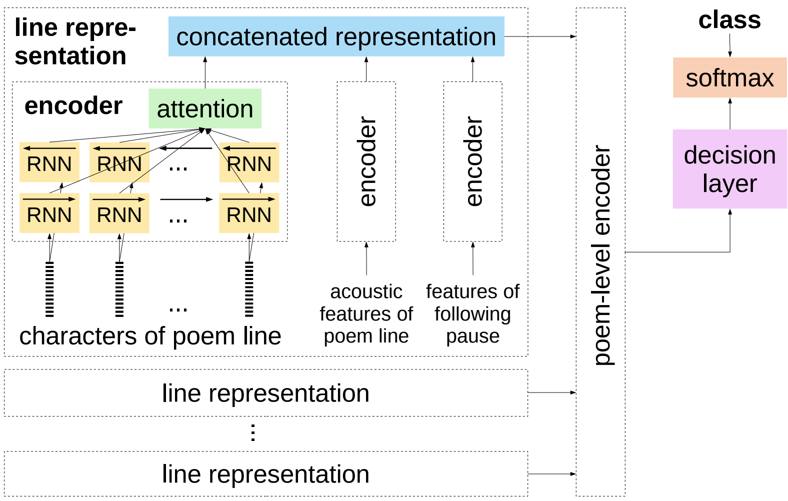

Figure 4: Full model for poetry style detection using neural networks.

Table 1: Classification results of the 6 rhythmical patterns using feature- and NN-based approaches.

classifier and feature engineering			NN and representation learning		
pause	parser	pause+parser	text	text+speech	text+speech+pause
0.36	0.37	0.47	0.47	0.66	0.73

4 Results

The results (weighted f-measure) of classifying poems in six classes by using feature- and NN-based approaches are presented in Table 1. As can be seen, the classifier using manually engineered features yielded the best results by using the pause and parser information (f-measure is 0.47). The classification results indicate that the method based on neural networks outperforms the manually engineered features, in particular when taking speech (and pausing) into account (f-measure is 0.73). This indicates that the neural network is better able to make use of information contained in the speech audio than can be captured by traditional feature-engineering approaches.

5 Conclusion and Future Works

We have captured the spectrum of free verse prosody in modern poetry using computational techniques along the fluency/disfluency continuum. We have trained classifiers that integrate for each line the textual information, the spoken recitation, as well as the pausing information, and integrate information across the lines within the poem by using feature- and NN-based approaches. We deem the overall classification performance as high (for NN-based approaches) (although classification is not our primary goal). In addition, we have trained classifiers for sub-problems and using reduced feature sets and the results obtained in these experiments support the expectations based on philological understanding. Our classifiers provide some insight into the decision making process via the attention mechanism and the possibility to map the internal state into lower-dimensional spaces for visualization. We find that our model indeed seems to internally re-create some notion of fluency. We also find that the mappings to lower dimensions do not fully support the claim of one single dimension of fluency.

In the future work, we intend to analyze the whole *lyrikline* corpus (and beyond) in order to gain insights about this broad sample of post-modern poetry. We hope to semi-automatically find additional poems that belong into one of our classes (and could be added as further training material after manual validation). Regarding the fact that enjambments are linked to syntactic characteristics, we could integrate syntactic information (such as part-of-speech (POS) tags) or try to pre-train our character encoding to decode the POS tag sequence. We also will inject further philological insight about what to pay attention into the training process in order to increase the philological validity of our model. With regards to the neural network, we could link the text and speech streams using connectionist temporal classification (Graves, Fernández, et al., 2006) for it to relate auditory to textual information in more detail, and we could connect the (sequential) line and pause encodings of audio for the model to better normalize out speaker specific (but not style-specific) characteristics.

Acknowledgements

The work is funded by the Volkswagen Foundation in the announcement ‘Mixed Methods in the humanities? Funding possibilities for the combination and the interaction of qualitative hermeneutical and digital methods’.

References

Baumann, T., Hussein, H., & Meyer-Sickendiek, B. (2018). Style detection for free verse poetry from text and speech. In *Proceedings of the 27th international conference on computational linguistics (coling 2018)*. Santa Fe, New-Mexico, USA.

Baumann, T., Köhn, A., & Hennig, F. (2018). The Spoken Wikipedia Corpus Collection: Harvesting, Alignment and an Application to Hyperlistening. *Language Resources and Evaluation*. doi:10.1007/s10579-017-9410-y

Berry, E. (1997). The free verse spectrum. *College English*, *59*(8), 873–897.

Beyers, C. (2001). *A History of Free Verse*. Fayetteville: University of Arkansas Press.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1724–1734). Doha, Qatar: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/D14-1179

Cooper, G. B. (1998). *Mysterious Music: Rhythm and Free Verse*. Stanford, CA: Stanford University Press.

Cureton, R. (1992). *Rhythmic Phrasing in English Verse*. Longman.

Finch, A. (2000). *The Ghost of Meter: Culture and Prosody in American Free Verse*. University of Michigan Press.

Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006). Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on machine learning* (pp. 369–376). ACM.

Graves, A. & Jaitly, N. (2014). Towards end-to-end speech recognition with recurrent neural networks. In *International conference on machine learning* (pp. 1764–1772).

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, *11*(1), 10–18. doi:10.1145/1656274.1656278

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., … Coates, A. et al. (2014). Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*.

Hussein, H., Meyer-Sickendiek, B., & Baumann, T. (2018a). Automatic Detection of Enjambment in German Readout Poetry. In *Proceedings of the speech prosody*. Poznań, Poland:

Hussein, H., Meyer-Sickendiek, B., & Baumann, T. (2018b). Tonality in Language: The "Generative Theory of Tonal Music" as a Framework for Prosodic Analysis of Poetry. In *Proceedings of the sixth international symposium on tonal aspects of languages*. Berlin, Germany:

Katsamanis, A., Black, M., Georgiou, P. G., Goldstein, L., & Narayanan, S. (2011). SailAlign: Robust Long Speech-Text Alignment. In *Proc. of workshop on new tools and methods for very-large scale phonetics research*.

Meyer-Sickendiek, B. (2012). *Lyrisches Gespür: Vom geheimen Sensorium moderner Poesie*. Fink, Wilhelm.

Rafferty, A. N. & Manning, C. D. (2008). Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines. In *Proceedings of the workshop on parsing german* (pp. 40–46). PaGe '08. Columbus, Ohio: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1621401.1621407

Silkin, J. (1997). *The Life of Metrical and Free Verse in Twentieth-Century Poetry*. Palgrave Macmillan UK.

Steele, T. (1990). *Missing Measures: Modern Poetry and the Revolt Against Meter*. Fayetteville: University of Arkansas Press.

Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., … Woelfel, J. (2004). *Sphinx-4: A Flexible Open Source Framework for Speech Recognition*. Mountain View, CA, USA: Sun Microsystems, Inc.

Wesling, D. (1996). *The Scissors of Meter: Grammetrics and Reading*. University of Michigan Press.