Analysis of Rhythmic Phrasing: Feature Engineering vs. Representation Learning for Classifying Readout Poetry

Timo Baumann
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, USA
tbaumann@cs.cmu.edu

Hussein Hussein and **Burkhard Meyer-Sickendiek**
Department of Literary Studies
Free University of Berlin
Berlin, Germany
{hussein,bumesi}@zedat.fu-berlin.de

Abstract

We show how to classify the phrasing of readout poems with the help of machine learning algorithms that use manually engineered features or automatically learnt representations. We investigate modern and postmodern poems from the webpage *lyrikline*, and focus on two exemplary rhythmical patterns in order to detect the rhythmic phrasing: The *Parlando* and the *Variable Foot*. These rhythmical patterns have been compared by using two important theoretical works: The *Generative Theory of Tonal Music* and the *Rhythmic Phrasing in English Verse*. Using both, we focus on a combination of four different features: The grouping structure, the metrical structure, the time-span-variation, and the prolongation in order to detect the rhythmic phrasing in the two rhythmical types. We use manually engineered features based on text-speech alignment and parsing for classification. We also train a neural network to learn its own representation based on text, speech and audio during pauses. The neural network outperforms manual feature engineering, reaching an f-measure of 0.85.

1 Literary Motivation and Introduction

Many theorists of modern poetry claim that accounts of meter touch on only a very limited part of the rhythmic structures and effects of modern and postmodern poems. For this reason, existing tools for the digital analysis of meter in poetry (Metricalizer (Bobenhausen, 2011)) do not capture the whole range of rhythmic features of modern poetry. Mainly the theory of Rhythmic Phrasing in English Verse (RPEV), developed by Richard Cureton, offers a detailed formalization of these rhythmic features, including a set of rules and a number of scanned examples. The RPEV draws heavily on music theory, mainly the *Generative Theory of Tonal Music* (GTTM), conceived by music theorist Fred Lerdahl and linguist Ray Jackendoff (Lerdahl and Jackendoff, 1983). Deeply influenced by Chomsky's generative grammar, they developed a musical grammar based on similar tree structure-style hierarchical organization uniting musical "phrase groupings". Such a grouping distinguishes the notion of phrases as relatively closed, self-contained musical units from that of the articulated phrasing associated with performance. An example of such a group is the *musical phrase*: "the smallest musical unit that conveys a more or less complete musical thought. Phrases vary in length and are terminated at a point of full or partial repose, which is called a cadence." (White, 1976, pp. 43-44).

Richard Cureton's theory on *Rhythmic Phrasing in English Verse* is truly the most important application of GTTM to poetry. This becomes evident in the hierarchical system used in both theories. The GTTM from Lerdahl and Jackendoff is based on four hierarchical systems that shape our musical intuitions: (1) The *Grouping structure* is based on the hierarchical segmentation of the musical piece into motives and phrases. (2) The *Metrical structure* identifies the regular alternation of strong and weak beats at a number of hierarchical levels, differing between the beat and the time span between two beats. Both 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. (4) The *Prolongational reduction* provides our "psychological" awareness of tensing and relaxing patterns in a given musical piece: In a strong prolongation, the roots, bass notes, and melodic notes are identical which effects the feeling

Proceedings of Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, pages 44–49
Santa Fe, New Mexico, USA, August 25, 2018.

of continuity and progression, caused by a movement towards relaxation. Following this hierarchical system by Lerdahl and Jackendorff, Richard Cureton has divided the poetic rhythm into three (not four) components: meter, grouping and prolongation (Cureton, 1992, pp. 124). The meter contains the perception of beats in regular patterns, the grouping refers to the linguistic units gathered around a single climax or peak of prominence, and the prolongation refers to the anticipation and overshooting of a goal, the experience of anticipation and arrival, such as the end of a line in an enjambment. Rhythmic prolongation is a matter of connected, goal-oriented motion, for example in the prosodic phrasing of an enjambment, where the line-break is felt as a linear extension of the sentence before the end of the sentence is reached in the next line.

Cureton's rhythm theory involves the interrelationship of these three components within a strictly hierarchical structure. A rhythm consists of a series of local events or units that are perceived as more or less prominent elements within longer events or units, which in turn are perceived as more or less prominent elements within even longer events or units, and so on to the entire poem. The analysis of phrase movements for Cureton involves examining the interaction of grouping and prolongation in a hierarchical organization. Cureton represents grouping hierarchies by a tree diagram (borrowed from linguistics) or an equivalent bracketing around each group. One of the examples he examines at length is a passage from W. C. Williams' poem Paterson. In *Paterson V* (1958) as well as in his late volumes *The Desert Music* (1954), and *Journey to Love* (1955), Williams developed the "triadic line," also known as the *Variable Foot*. It is based on the idea that, despite the different number of syllables per line, all the lines are isochronic, because all lines are based on a similar phrase/clause. In his readings, Williams emphasized the isochronicity of the lines by interrupting each by a regular breathing pause.

1.1 Applying Rhythmic Phrasing to Readout Poetry Analysis

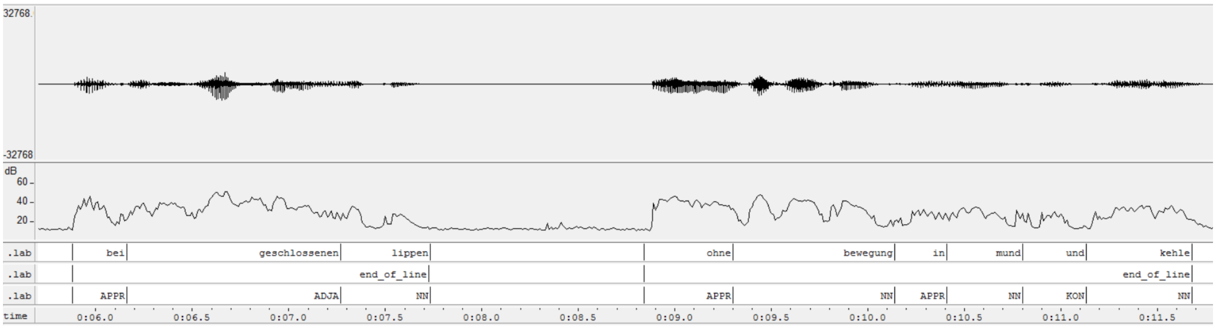
In our research, we analyzed a large number of German poems following this rhythmical type. One example is the following poem of Ernst Jandl – *Beschreibung eines Gedichts* (Jandl, 1982, pp. 129) – which uses the *Variable Foot* and is shown in Figure 1a:

bei geschlossenen lippen
ohne bewegung in mund und kehle
jedes einatmen und ausatmen
mit dem satz begleiten
langsam und ohne stimme gedacht
ich liebe dich
so daß jedes einziehen
der luft durch die nase
sich deckt mit diesem satz
jedes ausstoßen der luft durch die nase
das ruhige sich heben
und senken der brust

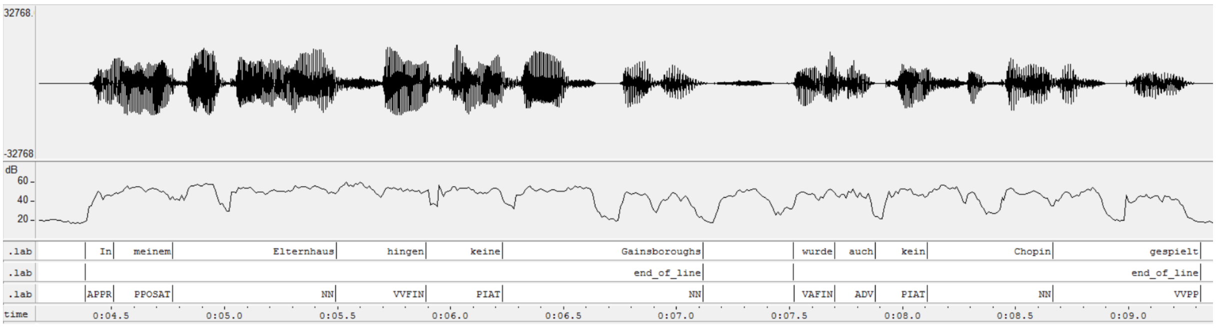
Jandl uses the *Variable Foot* and its breath-controlled line, which divides the syntax into a phrase or clause per line. That each line corresponds to exactly one single breath unit, causing a short break – a breathing space – at the end of each line, can be seen in Figure 1a: There is a characteristic gap at the end of the first line.

With regards to similar "phrase groupings" in modern and postmodern poetry, we compared the *Variable Foot* with a distinct but similar pattern, also using a sub-category below the sentence-level, that is a phrase/clause in each line. This second rhythmical pattern is called the *Parlando*, which was also very common in postwar German poetry. It was developed by the German poet Gottfried Benn. The *Parlando* is a prosodic style similar to the litany, using a similar orientation towards everyday speech in order to express the speaker's spontaneous feelings. A prominent example is Benn's poem "Teils-Teils" (Benn, 2006, pp. 317) which is shown in Figure 1b:

In meinem Elternhaus hingen keine Gainsboroughs
wurde auch kein Chopin gespielt



(a) *Variable Foot* pattern: Ernst Jandl's "beschreibung eines gedichtes" (English: description of a poem)



(b) *Parlando* pattern: Gottfried Benn's "TEILS-TEILS" (English: Half Here, Half There)

Figure 1: Two examples of the styles: poem text on the left, visualization of the first two lines on the right.

> ganz amusisches Gedankenleben
> mein Vater war einmal im Theater gewesen
> Anfang des Jahrhunderts
> Wildenbruchs »Haubenlerche«
> davon zehrten wir
> das war alles.

Both patterns – *Variable Foot* and *Parlando* – use a similar line arrangement, based on a colon in each line, as long as nearly each line has an enjambment: However, the *Parlando* makes no use of the breath-controlled line. Both patterns had a strong impact on German poetry beginning in the same period, the 1960s and 1970s. The exemplary analysis is particularly devoted to the GTTM, respectively to RPEV (Cureton, 1992) which is based on the GTTM. The GTTM and the RPEV both offer a very fruitful framework for the manual and digital analysis of these rhythmic patterns and for the specific "tonality" of (post-)modern poems. Given this theory, both poetic patterns use a similar line arrangement and a similar kind of prolongation, caused by the incomplete syntax at the end of nearly each line: the meaning runs over from one poetic line to the next. But in the *Parlando* style the poet does not emphasize the stops at the end of each line, in difference to those poets using the *Variable Foot* pattern. This can be clearly observed when listening to the audio recordings of both patterns.

With regards to the two patterns Cureton offered a new insight by "defining these line-terminal syntactic expectations as mid-level prolongational energies" (Cureton, 1992, pp. 153): Both patterns involve the experience of anticipating a goal at the end of each line, caused by the enjambment and its connection to the second part of the sentence in the following line. So both patterns use prolongation in nearly every line. But only the *Parlando* ignores this prolongation and its enjambment by arriving immediately at the goal in the next line. Only in the *Parlando*, the authors reading includes a time-span reduction.

1.2 Research Question and Hypothesis

We focus on structural similarities between tonality and cadences in music as well as poetic languages by using hermeneutical and computational methods. Our aim is to detect the tonality-like features of rhythmical patterns in a corpus of modern readout poetry and to use such features for classification.

Given that literary theory establishes contrastive features that differentiate the given styles (as outlined above and to be detailed below), we expect that we can automatically extract such features from the poems using language and speech processing tools and use them for classification. We contrast this approach to one where a hierarchical neural network (NN) learns its own representations based on the poetic source (text, speech, and pause between lines), rendering manual feature engineering and extraction unnecessary.

2 Database

In the project *Rhythmicalizer* (`www.rhythmicalizer.net`), we want to offer a theoretical as well as digital framework for the automatic recognition of rhythmical patterns in modern and postmodern poetry. We use a large collection of modern and postmodern readout poetry taken from our partner *lyrikline* (`www.lyrikline.org`) which hosts contemporary international poetry as audio files (read by the authors themselves) and texts (original versions & translations). The digital material covers more than $10,800$ poems by more than $1,200$ international poets from 80 different languages. This work investigates only poems written in German. The philological scholar (third author) in our project collected from the website poems written in German that belonged to either of the two patterns based on his experience in literary study and analysis. The total number of poems in this study is 68 from 24 poets (34 poems in each class). To deal with the low amounts of data, we use 10-fold cross-validation in the experiments reported below.

3 Classification Based on Manually Engineered Features

Our manually engineered features make use of a number of speech and text processing tools: We use a text-speech aligner (Baumann et al., 2018b), which implements a variation of the SailAlign algorithm (Katsamanis et al., 2011) to create an alignment of the written poems and spoken recordings in order to extract temporal features, in particular pauses. While overall the alignment coverage of the tool is quite high, we did not check the accuracy of the alignments.

On the textual side, we detected the syntactic features, in particular the words' Part-of-Speech (PoS), in order to identify those poems (*Parlando* and *Variable Foot*) using a "dismemberment of the line"(Berry, 1997, pp. 880) by separating the sentences into a nominal phrase and a verbal phrase. We use the Stanford parser (Rafferty and Manning, 2008) to parse the written text of poems, parsing each line in isolation. Poems are difficult material due to the absence of punctuation, special characters, and unexpected upper-/lowercasing which all introduce errors in the parsing process.

As could be seen in Figure 1, *Variable Foot* introduces longer pauses between lines. Different features including pause and parser information used in the classification process. Three feature sets are utilized: The **pause** feature set contains two features (the the average pause length at the end of each line as well as between words). Based on the parser output, we compute three features (the poem's number of lines, number of lines with a finite verb, and number of lines with punctuation) as the **parser** feature set. The **pause+parser** feature set includes five features which are a combination of 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) and settled for *AdaBoostM1* (Freund and Schapire, 1996) which yielded the best results (see (Hussein et al., 2018) for more details).

4 Neural Network-based Representation Learning

We train a neural network that learns to derive and represent features relevant for differentiating the patterns on its own. Inspired by Yang et al. (2016; Tsaptsinos (2017), we build a hierarchical attention network that encodes each line of a poem using a bidirectional recurrent network based on gated recurrent units (Cho et al., 2014) and *inner attention* (Liu et al., 2016). The result for each line is then combined by another bidirectional recurrent layer into a poem representation that is used for the final classification layer. Our model is implemented in *dyNet* (Neubig et al., 2017). Further details on the model are available in (Baumann et al., 2018a).

We use three variants for the input into the network: (a) only the text character sequence of each line, (b) we add acoustics of each line based on MFCCs and encoded similarly to the characters, and (c) we

Table 1: Results (weighted f-measure) for both approaches.

classifier and feature engineering			NN and representation learning		
pause	parser	pause+parser	text	text+speech	text+speech+pause
0.59	0.69	0.62	0.65	0.85	0.85

add acoustics of the *pause* following the line before the next line.

This method only requires a line-by-line text-audio alignment and no further text processing tools. While characters are not informative by themselves, the model is theoretically able to learn textual features such as the ones identified in manual feature engineering based on their character sequences.

5 Results and Discussion

In the *Parlando* subcorpus, we find per average 37 lines, 18 lines with finite verbs, and 25 lines using a punctuation. In the *Variable Foot* subcorpus, the same distribution is 20, 10, and 11. This indicates that the poetic lines in both classes do hardly contain complete sentences and that these poems belong to both classes: *Parlando* and *Variable Foot*. The results of classifying poems as dominated by *Parlando* or *Variable Foot* are presented in Table 1. As can be seen, the classifier using manually engineered features yielded the best results by using only the parser information (f-measure is 0.69) which is unexpected given that pauses identify the *Variable Foot* pattern, according to theory. 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. The NN that uses only text is inferior to manual parsing features. 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.

6 Conclusion and Future Work

We presented an experiment for the classification of rhythmical patterns in modern and postmodern poetry by analyzing a corpus of readout poems using machine learning techniques (using manually engineered feature or representation learning). We compared these rhythmical features with rhythmic phrasing in readout poetry and focused on two important rhythmical patterns (*Parlando* and *Variable Foot*). We used different sets of manually engineered features based on pause and parser information for classification. We found that parser features outperform pause features, although the latter should have been favored based on theoretical insight. Furthermore, we find that the feature-less neural networks-based approach outperforms the methods based on manually derived features. This indicates that elaborate feature engineering can be offset by representation learning capabilities of neural networks.

In the future, we hope to understand better the *aspects* of a poem that are encoded in its representation. A total of 18 rhythmical patterns are defined till now by the philological scholar. We want to analyze and classified other rhythmical patterns.

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

Timo Baumann, Hussein Hussein, and Burkhard Meyer-Sickendiek. 2018a. 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.

Timo Baumann, Arne Köhn, and Felix Hennig. 2018b. The Spoken Wikipedia Corpus Collection: Harvesting, Alignment and an Application to Hyperlistening. *Language Resources and Evaluation*.

Gottfried Benn. 2006. *Sämtliche Gedichte / Künstlerische Prosa - Band 1*. Klett-Cotta, Stuttgart.

Eleanor Berry. 1997. The Free Verse Spectrum. *College English*, 8(59):873–897.

K. Bobenhausen. 2011. The Metricalizer – Automated Metrical Markup of German Poetry. In C. Küper, editor, *Current Trends in Metrical Analysis*, pages 119—131.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 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)*, pages 1724–1734, Doha, Qatar, October. Association for Computational Linguistics.

R.D. Cureton. 1992. *Rhythmic Phrasing in English Verse*. Longman.

Yoav Freund and Robert E. Schapire. 1996. Experiments with a New Boosting Algorithm. In *Thirteenth International Conference on Machine Learning*, pages 148–156, San Francisco. Morgan Kaufmann.

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

H. Hussein, B. Meyer-Sickendiek, and T. Baumann. 2018. 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, June.

Ernst Jandl. 1982. *der gelbe Hund. Gedichte*. Luchterhand Verlag, Darmstadt und Neuwied.

Athanasios Katsamanis, Matthew Black, Panayiotis G Georgiou, Louis Goldstein, and S Narayanan. 2011. SailAlign: Robust Long Speech-Text Alignment. In *Proc. of Workshop on New Tools and Methods for Very-Large Scale Phonetics Research*.

Fred Lerdahl and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press.

Yang Liu, Chengjie Sun, Lei Lin, and Xiaolong Wang. 2016. Learning natural language inference using bidirectional LSTM model and inner-attention. *CoRR*, abs/1605.09090.

Graham Neubig, Chris Dyer, Yoav Goldberg, Austin Matthews, Waleed Ammar, Antonios Anastasopoulos, Miguel Ballesteros, David Chiang, Daniel Clothiaux, Trevor Cohn, Kevin Duh, Manaal Faruqui, Cynthia Gan, Dan Garrette, Yangfeng Ji, Lingpeng Kong, Adhiguna Kuncoro, Gaurav Kumar, Chaitanya Malaviya, Paul Michel, Yusuke Oda, Matthew Richardson, Naomi Saphra, Swabha Swayamdipta, and Pengcheng Yin. 2017. Dynet: The dynamic neural network toolkit. *arXiv preprint arXiv:1701.03980*.

Anna N. Rafferty and Christopher D. Manning. 2008. Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines. In *Proceedings of the Workshop on Parsing German*, PaGe ’08, pages 40–46, Stroudsburg, PA, USA. Association for Computational Linguistics.

Alexandros Tsaptsinos. 2017. Lyrics-based music genre classification using a hierarchical attention network. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, pages 694–701.

John D. White. 1976. *The Analysis of Music*. New York: Prentice-Hall, Inc.,.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489.