Deep Learning meets Post-modern Poetry[*]

Timo Baumann[1][0000−0003−2203−1783] and Burkhard Meyer-Sickendiek[2]

[1] Department of Informatics, Universitt Hamburg, Germany
baumann@informatik.uni-hamburg.de
[2] Department of Literary Studies, Freie Universitt Berlin, Germany
bumesi@zedat.fu-berlin.de

http://www.rhythmicalizer.net

Abstract. We summarize our project *Rhythmicalizer* in which we analyze a corpus of post-modern poetry in a combination of qualitative hermeneutical and computational methods, as we have run the project over the course of the past three years (and preparing it for some time before that). Interdisciplinary work is always challenging and we here focus on some of the highlights of our collaboration.

Keywords: Literary Studies · Machine Learning · Meta-Research.

1 Introduction

At least 80 % of modern and post-modern poems exhibit neither rhyme nor metrical schemes like iamb or trochee. However, does this mean that they are free of any rhythmical features? Of course not and the US American research on *free verse prosody* claims the opposite: Modern poets like Whitman, the Imagists, the Beat poets as well as contemporary Slam poets have developed a *post-metrical* idea of prosody, using rhythmical features of everyday language, prose, and musical styles like Jazz or Hip Hop, yielding a large and complex variety in their poetic prosodies which, however, appear to be much harder to quantify and regularize than traditional patterns.

In our joint project, we examine the largest portal for spoken poetry *Lyrikline*[3] and analyze and classify such rhythmical patterns in a *human-in-the-loop* approach in which we interleave manual annotation with computational modelling and data-based analysis [5].

The remainder of this paper is structured as follows: in the next section, we describe the research questions that we set out to address in our project; in Sections 3 and 4, we describe the methods used and the results obtained, respectively, and in Section 5 we describe – as much as is adequate in a public setting – the lessons *taken* during the course of the project. (It is still unclear, if all of these lessons also were lessons learned.)

---

[3] http://lyrikline.org

2   Research Questions

Automating the analysis and differentiating the prosodies of free verse post-modern poetry comes with a multitude of challenges, including the manual analysis and classification of sufficiently many poems as training material for an automated method. Yet, the amount of data that can be acquired (even in the order of hundreds of poems) is still very little for modern machine learning methods. However, the by far largest challenge of this endeavour lies in the nature of free verse prosody, which is an almost purely *spoken* aspect of a poem and is not readily observable from the pure textual form as in traditional metric schemes (see, e.g., [6]). We thus base our analyses on the audible form of the poem, spoken aloud by the original author, as a gold standard for the intended prosodic realization of the poem (instead of, e.g., attempting to derive such features in a generic way from the written form). The use of speech data greatly improves our modelling ability, yet it also increases the complexity of the data and the need for automated analyses.

Our primary research questions are hence: (a) can we relate theories on the classification of the prosodies found in post-modern poetry to poems that we find our collection of spoken poems, (b) can we pre-process the poems in our collections in such a way that they can be automatically analyzed, (c) can we build an automatic classification system that uses quantitative features derived from the poems to yield the classification despite the significant *data sparsity*, and (d) can we gather additional, new insight from the automatic classification in the humanities.

3   Method

We here describe the setup of our project, in particular how we went about defining and implementing the philological classification, as well as the processes used to prepare and automatically classify our data. Over the course of the project, we have extended and sometimes changed our methods, as is reflected in our publications, and we here only describe a subset of the methods used.

3.1   Theoretical Foundation: Grammetrical Ranking and Rhymic Phrasing

Our classification of poetry is based on two theoretical approaches: 1) the Idea of grammetrical ranking and 2) the Idea of rhythmic phrasing. 1) 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. [10] 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: "Grammetrics assumes that meter and grammarcan be scissored by
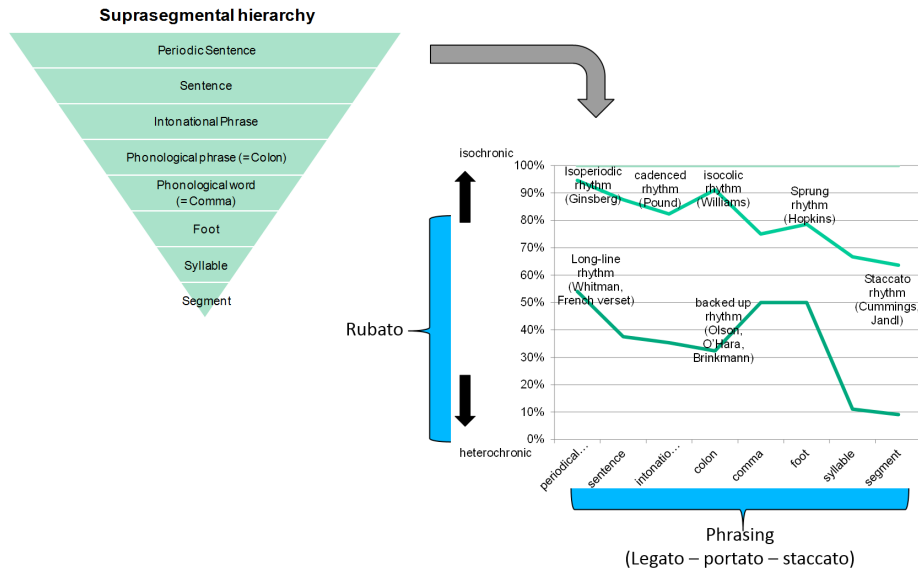
Fig. 1. Working hypothesis for a prosodic hierarchy for *free verse poetry*, as well as a placement of poetic styles along the two axes governing unit and regularity of temporal arrangement.

each other, that the cutting places can be graphed with some precision. One blade of the shears is meter, the other grammar. When they work against each other, they divide the poem. It is their purpose and necessity to work against each other." [10]

2) The concept of rhythmic phrasing was developed by Richard Cureton. For Cureton, rhythm embraces what has traditionally been regarded as very different kinds of perceptual phenomena. Cureton divided rhythm - a global term covering all relations of strength and weakness - into three distinct components, which he terms meter, grouping, and prolongation. Meter involves the perception of beats in regular patterns; grouping involves the apprehension of linguistic units organized around a single peak of prominence; and prolongation involves the experience of anticipation and arrival. Curetons claimed a hierarchical, multi-dimensional, and preferential treatment of poetic rhythm, going back to [9] treatment of rhythm in Tonal Music. Lerdahl and Jackendoff claimed four different rhythmic dimensions: a) The grouping structure in terms of a hierarchical segmentation, b) the metrical structure in terms of a regular alternation of strong and weak beats at a number of hierarchical levels, c) the time span-reduction in terms of an organizations uniting time-spans at all temporal levels of a work, and d) the prologational reduction in terms of a 'psychological' awareness of tensing and relaxing patterns in a given piece.

As a result, the prosodic hierarchy to be considered for free verse is considerably more complex than for metric poetry and our working hypothesis of this

hierarchy is depicted in Figure 1 (left). As can be seen, all levels of the linguistic hierarchy can carry poetic prosodic meaning, from the segment up to the periodic sentence. Based on this prosodic hierarchy, Figure 1 (right) depicts a categorization of some poets' works along two axes, the governing prosodic unit (x-axis) and the degree of iso/heterochronicity, or regularity of temporal arrangement (y-axis). The respective rhythm derives from the combined prosodic units and the degree of its isochronic (or heterochronic) succession. The green lines are meant as a localization of poets according to the interplay of grammetrical ranking and prosodic succession. The blue brackets mark the time span reduction (Rubato) as well as the prolongational reduction (Phrasing). According to the idea of time-span-reduction, the rubato is caused by a deviation from the isochronous rhythm. And according to the idea of prolongational reduction, the phrasing is divided into three different articulation techniques (legato, portato, staccato).

3.2   Manual Analysis and Annotation

Based on our theoretical foundation, we devised a number of prosodic classes of poems and build a manually curated collection of poems for each class. Over the course of our project, we extended the kinds of poems to differentiate, starting with only various kinds of sound poetry (letristic vs. syllabic decompositions) via kinds of poems that differ by the way that enjambments are realized (variable foot poems vs. unemphasized enjambments vs. gestic rhythms) to a large collection of 18 classes. Presently, we work on identifying and specifying the relation of the classes to each other which will help inform the automatic classification methods described below.

3.3   Data Extraction and Preparation

We work with the website *Lyrikline* which hosts a large collection of modern and post-modern readout poetry. *Lyrikline* was created by the Literaturwerkstatt Berlin and hosts contemporary international poetry as audio files (read by the authors) and texts (original versions & translations), so it offers the melodies, sounds, and rhythms of international poetry, recited by the authors themselves. It covers more than 10,000 poems from about 1000 international poets from more than 60 different countries. Nearly 80 % of these are postmetrical poems. We focus our analysis on the roughly 2400 German-language poems on the page.[4]

To enable our analyses, we perform forced alignment [8] (which we refine manually as necessary) in order to be able to relate speech and text line-by-line. In addition, we use various tools to extract syntactic information from the text as required.

---

[4] An extraction program for extracting poems (text, audio and meta-data) is available by request from the first author.
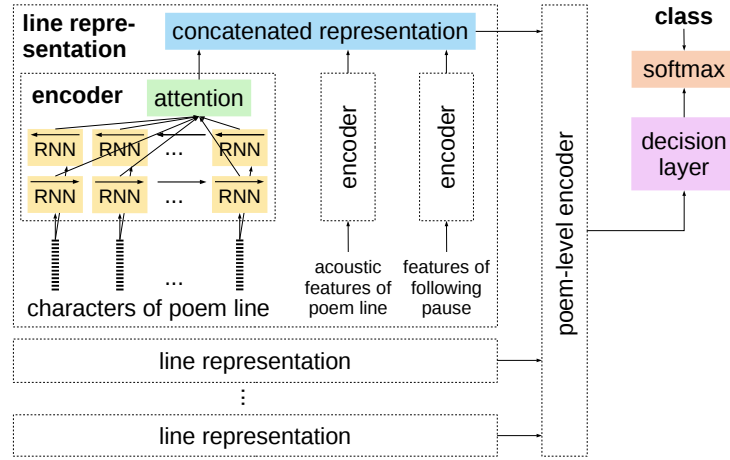
Fig. 2. Full model for poetry style detection: each line is encoded character-by-character by a recurrent neural network (using GRU cells) with attention. Acoustic features of each line, as well as of the pause following up to the next line, are encoded similarly. Per-line representations are concatenated and passed to a poem-level encoder. The final decision layer optimizes for the poem's class.

3.4   Classification Methods

We devised our project with primary concerns on *data sparsity* (i.e., the limitation that there is too little data to fully train advanced machine learning algorithms) as well as *interpretability* (i.e., the ability of a method to explain what aspects of the data determine its behaviour).

Thus, we first focused our efforts on building interpretable classifiers (e.g., decision trees) on specific interpretable features (e.g., the presence or absence of a verb in a line). However, based on the (relatively obvious) fact that the prosodic structure of a poem is reflected in most of its lines, and that our interpretable features need to be aggregated across the poem's lines anyway, we also build a hierarchical model for our poems, as depicted in Figure 2.

Our deep-learning based hierarchical attention model [11] uses multiple recurrent neural-network-based encoders for each line, focusing on the text, the speech, and the quality of the pause following the line, respectively, and using inner-attention. We aggregate across the lines with another recurrent layer. The architecture is more thoroughly described in [1]. Given the fact that most lines exhibit the structural properties that yield a poem's prosodic classification, we use a two-stage training procedure, in which we first train the line-by-line classification in isolation and only afterwards train the full network including the final decision layer.

4   Results

Over the course of our project, we have performed a number of experiments; we here only give a very rough overview of the results and refer the reader to the corresponding publications for further detail.

In [1], we described our first experiments on classifying poetry using a hierarchical attention network. We found that free verse prosodies can be classified along a *fluency continuum* and that the classifier's mis-classifications cluster along this fluency continuum. Via ablation studies (i.e., leaving out certain features to find their relative importance for the final results), we found that, depending on the classes analyzed, not only is the acoustic realization of the speech itself highly relevant for classification but also the realization of the *pause* following each line of the poem. In fact, our classifier is highly reliable in determining different kinds of enjambments [3] and that an additional annotation of enjambments in the poems is unnecessary (unlike what we had originally hypothesized). We also tried traditional classification approaches using engineered features [7] and in a comparison with our deep learning-based method found these to be inferior [2], although they have direct access on theoretically grounded features. Recently, we work on integrating such symbolic features into our deep learning-based approach.

We find that pre-training and data augmentation are crucial for the success of neural learning techniques on the (relatively) small data sets typically found in digital humanities. Another way to success on difficult-to-describe material like sound poetry is the granularity of modelling: instead of using words as the textual unit of processing, our recurrent networks use the sequence of individual characters found in the poem. This has (at least) the following two advantages: there is a fixed set of characters (as compared to the endless possibilities for words) which means that our models has far fewer parameters to train and cannot be confronted with 'out-of-vocabulary words' when applied to a new poem. Secondly, the concept of 'word' does not necessarily reflect what we find in sound poetry and the prosodic features are actually very well reflected by the stream of characters (e.g., the recurrence of consonant-vowel pairs) as compared to words.

Finally, and coming back to our original ideas [5], we have implemented an interface for *humanist-in-the-loop* classification and analysis of our corpus [4] and have gradually applied our method to eventually cover the full corpus.

5   Lessons (Learned?)

Over the course of our interdisciplinary project, the common research question has been a strong uniting factor given the sometimes differing research interests involved (creating classification systems that work despite data sparsity vs. the interest in finding evidence in data for humanistic theories). A strong enabling factor in the joint work was the ability to openly *ask* about the other field's theory, applications, limitations, and to not be shy to ask again if the given answer was unclear. It often helped to *agree to disagree* and to move on despite of this, or to compromise on what should be done.

The project sometimes was hindered by technical limitations, such as forced alignment of text to speech (or syntactic analysis) breaking down for sound poetry, and often being too inaccurate (or with too low coverage) for large amounts of the material. This meant that a great deal of manual annotation of the *base* material had to be performed, despite the later stages being successfully automated. This means that the classifiers produced in the project cannot always be applied to new data in a fully automatic way and the work required to prepare data (although relatively straightforward) can be more than would be required for a manual prosodic analysis itself.

References

1. Baumann, T., Hussein, H., Meyer-Sickendiek, B.: 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 (2018)
2. Baumann, T., Hussein, H., Burkhard: Analysis of rhythmic phrasing: Feature engineering vs. representation learning for classifying readout poetry. In: Proceedings of the Joint LaTeCHCLfL Workshop. pp. 44–49. Association for Computational Linguistics, Santa Fe, USA (sep 2018), `https://aclanthology.info/papers/W18-4505/w18-4505`
3. Baumann, T., Hussein, H., Meyer-Sickendiek, B.: Analysing the focus of a hierarchical attention network: The importance of enjambments when classifying post-modern poetry. In: Proceedings of Interspeech. pp. 2162–2166. Hyderabad, India (Sep 2018). https://doi.org/10.21437/Interspeech.2018-2533
4. Baumann, T., Hussein, H., Meyer-Sickendiek, B., Elbeshausen, J.: A tool for human-in-the-loop analysis and exploration of (not only) prosodic classifications for post-modern poetry. In: Proceedings of INF-DH. Gesellschaft fr Informatik, Kassel, Germany (sep 2019)
5. Baumann, T., Meyer-Sickendiek, B.: Large-scale analysis of spoken free-verse poetry. In: Proceedings of Language Technology Resources and Tools for Digital Humanities (LT4DH). Osaka, Japan (Dec 2016)
6. Bobenhausen, K.: The metricalizer2–automated metrical markup of german poetry. Current Trends in Metrical Analysis, Bern: Peter Lang pp. 119–131 (2011)
7. Hussein, H., Meyer-Sickendiek, B., Baumann, T.: Automatic detection of enjambment in german readout poetry. In: Proceedings of Speech Prosody. Poznn, Poland (jun 2018). https://doi.org/10.21437/SpeechProsody.2018-67
8. Katsamanis, A., Black, M., Georgiou, P.G., Goldstein, L., Narayanan, S.: SailAlign: Robust long speech-text alignment. In: Proc. of Workshop on New Tools and Methods for Very-Large Scale Phonetics Research (2011)
9. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music. MIT Press series on cognitive theory and mental representation, MIT Press (1983)
10. Wesling, D.: The Scissors of Meter: Grammetrics and Reading. University of Michigan Press (1996)
11. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: 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. pp. 1480–1489 (2016)