Large-scale Analysis of Spoken Free-verse Poetry

Timo Baumann
Department of Informatics
Universität Hamburg
baumann@informatik.uni-hamburg.de

Burkhard Meyer-Sickendiek
Department of Literary Studies
Freie Universität Berlin
bumesi@zedat.fu-berlin.de

Abstract

Most modern and post-modern poems have developed a post-metrical idea of lyrical prosody that employs rhythmical features of everyday language and prose instead of a strict adherence to rhyme and metrical schemes. This development is subsumed under the term *free verse prosody*. We present our methodology for the large-scale analysis of modern and post-modern poetry in both their written form and as spoken aloud by the author. We employ language processing tools to align text and speech, to generate a null-model of how the poem would be spoken by a naïve reader, and to extract contrastive prosodic features used by the poet. On these, we intend to build our model of free verse prosody, which will help to understand, differentiate and relate the different styles of free verse poetry. We plan to use our processing scheme on large amounts of data to iteratively build models of styles, to validate and guide manual style annotation, to identify further rhythmical categories, and ultimately to broaden our understanding of free verse poetry. In this paper, we report on a proof-of-concept of our methodology using smaller amounts of poems and a limited set of features. We find that our methodology helps to extract differentiating features in the authors' speech that can be explained by philological insight. Thus, our automatic method helps to guide the literary analysis and this in turn helps to improve our computational models.

1 Introduction

Lyrical analyses of poetry rely mainly on the poems' textual form, focusing on the analyst's philological insight of how to properly read a poem. Classic poetry has been analyzed extensively in this way, leading to a deep understanding of its prosodic structure which comprises rhyme and metrical schemes such as iambic or trochaic meter. The large amounts of manually analyzed works of such poetry have lead to tools like Metricalizer (Bobenhausen, 2011) which proposes metrical patterns given a poem's text and Sparsar (Delmonte and Prati, 2014) which uses such patterns for speech synthesis of metric poetry. These tools, however, do not work for *free verse poetry* which was started by modern and post-modern poets like Whitman, the Imagists, the Beat poets, and contemporary Slam poets. Regarding this kind of poetry, Finch (2000), Berry (1997), Silkin (1997), Meyer-Sickendiek (2012), Lüdtke et al. (2014) and many others manually analyze the prosodic forms and styles of some poems in great detail, providing a narrow but detailed view into free verse prosody. It will, however, not be possible to achieve a large phenomenal and analytical coverage by manual work alone.

At the same time, original recordings of modern and post-modern poets reciting their poetry are available, but neglected in philological research so far. We set out to change this. Through a collaboration with *Lyrikline*[1], we are able to use their speech and text database of modern and contemporary poetry, giving us access to hundreds of hours of author-spoken poetry. We aim to collaborate with further sources such as *PennSound* and *Poetry Foundation*. For the spoken and written poems, we create a text-speech alignment and using this alignment, we can extract a wide range of prosodic features as well as textual

[1]Lyrikline, `http://www.lyrikline.org` is an international website devoted to spoken contemporary poetry, established by *Literaturwerkstatt Berlin*, Germany.

features using various tools. We then plan to use machine learning to learn styles and rhythmical/prosodic figures (such as syncopation or cadence) based on annotations, to cluster similar poems (and poets) and their styles, or to identify 'outlier poems' which deserve further analysis.

In an iterative fashion, the results of automatic analyses are to be presented to a human philological analyst in a visual and understandable form and the interface to be developed will allow the annotation of particularities, including the addition of markables to annotate newly found types of noteworthy information. These will then be fed back to the machine-learning back-end and be used in the next cycle of automatic analyses. This *human-in-the-loop* approach to poetry analysis combines the strengths of human and machine analyses, namely deep understanding and broad coverage.

In the remainder of this paper we will first describe what sets apart free verse from the more traditional metric poetry in Section 2. This will highlight why traditional meter analysis tools cannot be used for free verse and why standard speech processing tools are more suitable for the task. In Section 3 we describe our methodology, in particular the contrasting of a null model to amplify the particularities of poetic speech. We describe our implementation in Section 4 and describe our preliminary experiments and results in Section 5. We close with a dicussion in Section 6 and outline future work.

2 Free Verse Poetry and Free Verse Prosody

The most important development in modern and postmodern poetry is the replacement of traditional meter by new rhythmical features: A structure of lyrical language was developed that renounced traditional forms like rhyme and meter, developing novel forms of prosody, accent, rhythm, and intonation to replace the traditional, and to forge a poetry instead based on the rhythms of contemporary American Speech (Gates, 1985; Gates, 1987). Music was another important influence, especially jazz, as well as efforts to visually register distinct free verse prosodies in print (Perloff, 1983). Prosody, as a specifically literary rhythm, was thus crucially redefined in modern American and European poetry.

This new kind of free verse prosody is marked by a new interplay of line and stanza, which may vary in different ways – line length, line integrity, line grouping, the dismemberment of the line, or systematic enjambement. William Carlos Williams developed the *isocolic* 'step-down line', a triadic alignment of tercets in which every line has its own arc of accentuation, while the gap between the lines is always the same size, resulting in a flowing rhythm. This mode of structuring a poem into cola (the rhethorical figure consisting of a clause which is grammatically, but not logically, complete) relates well to the idea of shallow parsing, or 'parsing by chunks' developed in linguistics (Abney, 1991). Under this influence, American Beat poet Allen Ginsberg develops an *isoperiodic* rhythm, in which lines are structured by "breath units", making them even closer to natural and fluent speech. Being influenced by Williams and Ginsberg, the famous Black Mountain poet Charles Olson based his idea of the "projective verse" on a similar relationship between the line and the poet's breathing (Olson, 1966), now combining isochronic and isoperiodic lines in order to create a more heterochronic rhythm (Golding, 1981). Even below the syllable, *sound poets* like Ernst Jandl used a prosody based on individual rhythmic phonemes.

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 (a). 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 (b) 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), according to philological analysis (Lüdtke et al., 2014). The green lines are meant as an estimate that outlines the limits of free verse poetry to prose.

To summarize, modern and post-modern free verse poetry has developed a broad range of prosodic styles stemming from a diverse set of structuring principles. Its prosody is often guided more by everyday speech than the prosodic principles of traditional poetry.

3 Methodology

Our methodology is to employ automatic analyses based on computational speech and language processing in combination with manual hermeneutical analyses. We will use machine learning techniques in a *human-*

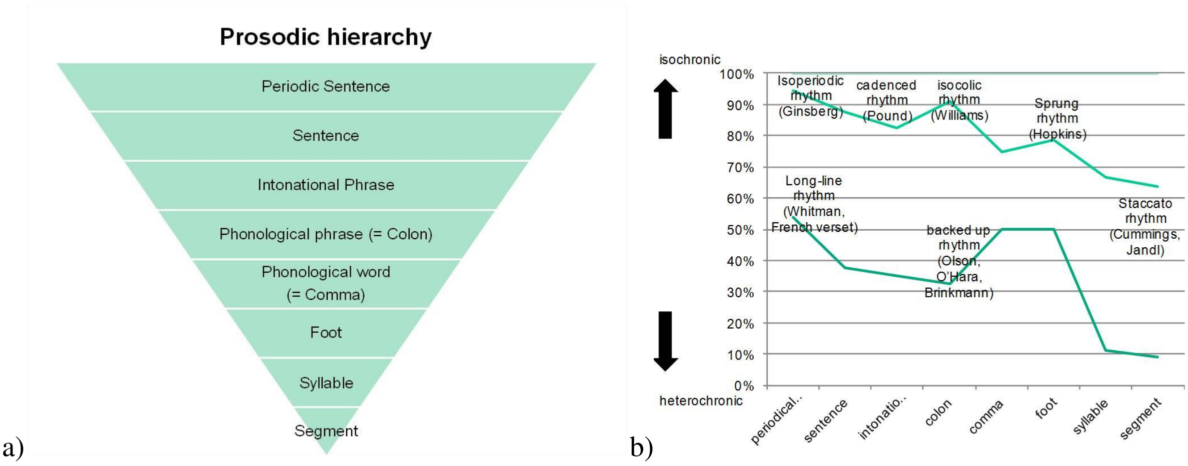

Figure 1: (a) Working hypothesis for a prosodic hierarchy for free verse poetry, as well as (b) a placement of poetic styles along the two axes governing unit and regularity of temporal arrangement.

in-the-loop approach, in which we cycle between building (or extending) computational models and manual philological analysis of phenomena (and the annotation of these phenomena).

The automatic analysis will, of course, be based on automatically extracted features that are potentially useful to describe and differentiate free verse prosody. Our goal is to find exactly those features and combinations that constitute free verse prosody. A prerequisite for computing such features is to create a text-speech alignment for the written poems and spoken recordings. In a next step, we extract phonological prosodic features such as ToBI labels (Silverman et al., 1992), which form the basis for phonetic prosodic features such as TILT (Taylor, 1998) or PaIntE (Möhler, 2001) parameters that describe individual tones, duration and loudness features, as well as silences as a basis for rhythmical structure and the level of isochronicity.

In order to build a *poetic prosody model* rather than just a prosody model for poetry, we contrast the features that we extract from the authentically spoken poems against a poetically naïve automatic reading. This *null model* of read-out poetry helps us to accentuate the peculiarities of the specific poetic styles and reduces the negative impact of data sparsity by focusing our analysis on the outstanding aspects of a poem.

Regarding higher-level prosodic analysis and feature extraction, poems will be split into prosodic segments and the combination of these prosodic segments will be assigned to types of rhythm. The patterns remain to be developed, since the existing research only discusses certain "figures of sound" or "figures of rhythm", both of which involve the repetition of some key linguistic component (Cooper, 1998). This processing step may yield similar results as depicted in Figure 1 (b) above. In addition, we use unsupervised learning, such as clustering techniques and outlier detection in order to steer the manual philological analysis towards potentially interesting parts and phenomena in the large corpora. We use existing meta-information (e.g. about poetic type) to train classifiers.

In our procedure, we will analyze classifier models with the aim of generating explanations for poetic categorization (e. g. RIPPER (Cohen, 1995) induces rule-based models that are easy to analyze), and these explanations can be valued (and in the strongest case rejected) by the human analyst. In this way, the human expert is able to steer the prosody modeling process away from computationally optimal but philologically ungrounded decision-making towards those aspects of poetic prosody that are deemed philologically relevant.

4 Implementation

We use text-speech alignment software as presented by Köhn et al. (2016) using a variation of the SailAlign algorithm (Katsamanis et al., 2011) based on the Sphinx-4 speech recognizer (Walker et al., 2004). Our prosodic analyses so far are limited to ToBI and we use AuToBI (Rosenberg, 2010) to generate

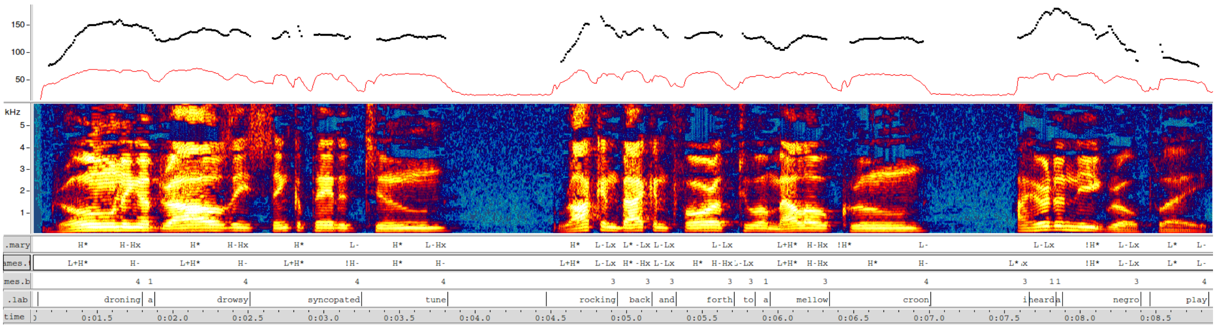

Figure 2: Example analysis of the first lines of Langston Hughes: “The Weary Blues.” Shown from bottom to top: time, word alignment, ToBI breaks and tones, null-model estimate of ToBI tones, spectrogram, signal power (dB) and pitch (in Hz).

an automatic intonation annotation.

To generate our null-model version of a naïve, non-poetic reading of the poem’s text, we use MaryTTS (Schröder and Trouvain, 2003) to synthesize speech audio which we then feed into the same alignment and analysis pipeline as we use for the author’s speech. In this way, we limit the effect of MaryTTS and AuToBI internally using different interpretations of the ToBI standard.

In our preliminary experiments, we perform simple comparisons of the relative and contrastive occurrence of ToBI labels in read-out poems of different styles. Once annotations of notable structures become available, machine-learning tools such as WEKA (Witten and Frank, 2000) will be used as outlined above.

Our system is planned to be developed into a web-based client and server architecture. In this way, the human-in-the-loop hermeneutical analysis and interaction can be performed via any computer (and annotations can be parallelized). Furthermore, there is no need to install and update any software on the client-side, minimizing the risk of inconsistent data-handling or versioning issues.

5 Preliminary Experiments

We have performed preliminary experiments based on 10 poems from different modern and postmodern poets that cover the full range of free verse poetry and using only a limited set of prosodic features.

Regarding automatic text-speech alignment, we find positive results. For half of the poems, 90-98 % of the words after tokenization are successfully forced-aligned, which we deem a sufficient quantity to perform further analyses. Particular outliers in our sample with only 15 % of words aligned are W.C. Williams (very old recordings and a softly mumbling voice) as well as A. Stewart, who’s experimental permutative poetry is overlaid with music and heavy echos, and C. Bök, who’s segmental sound poetry already defies the trained grapheme-to-phoneme conversion models. Alignment of the (synthesized) null-model data is near perfect, as expected. Alignment quality seems reasonable but we have not yet formally evaluated it (e.g. in terms of root-mean-squared-error – RMSE – of boundary placement).

Regarding prosodic modelling, we have focused on ToBI labels so far, as these are less prone to pick up speaker characteristics than phonetically motivated features such as TILT parameters. AuToBI successfully produces tone alignments for aligned poetry (including the null-model audio), and again, although we have not formally evaluated the quality, annotations appear reasonable. However, the particularities of poetic speech are not covered in AuToBI’s standard models and one future goal is to improve or extend these models over the course of our project. An example analysis is presented in Figure 2.

By contrasting the ToBI labels found in the author’s speech with those for the null model, we can amplify what makes the prosodic style of a poem special. We find that in the first line, all `L+H*` accents correspond to `H*` accents in the null model, the passage *back and forth* receives two `H*`s instead of just one `L*`. Such an accumulation of differences struck the first author (a relatively theory-agnostic speech scientist) to be somewhat notable, and the second author (the philological expert on the team) explained that precisely this could be one manifestation of the poetic *syncopation* appearing in the poem.[2] Our

[2]Notice that rhythmical aspects such as low syllable rate in the first line contrasted by high syllable rate in the second are

human-in-the-loop approach works exactly like this: the system finds candidate peculiarities in poems based on a measure of surprisal, which can then be named, described and explained by the human expert. In this case, we decided to next annotate several syncopations in multiple poems to build a model of the syncopation phenomenon. This model will then be used to (a) find further syncopation candidates which can be assessed by a human annotator in order to actively learn better models, and (b) as additional input into the style models to be developed.

The manual observations from the first line of the poem can be generalized by contrasting the occurrence of ToBI tones in the author's speech relative to the null model in the full poem. Using such statistics, we find that Hughes uses more than 3 times as many `L+H*` as the null model would. Looking at the confusion matrix (i.e., what Hughes uses instead of the tones expected by the null model), we find that these accents generally occur in places where the null model would use `H*` or no accent at all. We can compare different poems (and their prosodic styles) by comparing the respective differences to their null models.

We have performed a preliminary comparison based on tone differences of the aligned poems. We find that different deviations from the null model (as described above) occur in poems of different style. At least some of these differences do not appear to be based on speaker characteristics but on poetic style – like the `L+H*` which may be an artifact of the slow and connected speaking style interleaved with faster syncopations as in Hughes' example. We plan to next train classifying models for different poetic styles. However, the amount of data and features exploited so far is clearly insufficient to report even preliminary results of machine-learned models.

6 Discussion and Future Work

We have presented our procedure for the large-scale analysis of spoken free verse poetry. The prosody of free verse poetry in many cases uses the rhythms of everyday speech (with some exceptions and extensions), which is why we base our procedures on conventional speech processing tools. Our methodology aims to single out what makes a poem special and to build a model of prosodic styles based on these specialities. We use a *human-in-the-loop* approach which allows us to analyze large amounts of spoken poetry and to focus manual analysis on the most important aspects found in the corpus.

In our preliminary experiments, we found that robust forced alignment works reasonably well for spoken poetry, yet still poses some interesting research problems, such as emphatic speech, onomatopoeic expressions, consonant or vowel clusters, prosodic disfluencies caused by line-breaks, etc. Likewise, we believe that current tools for intonation analysis (such as AuToBI) can fruitfully be used to analyze free verse, but that adaptations are necessary to leverage the full potential.

Our initial experiments are limited in data but will be scaled up to cover hundreds of hours of poems, enabling the reasonable use of machine-learning techniques over a multitude of features, finally also including text processing methods. Poetry is a particularly interesting form of language as, being art, it does not focus on function but uses creativity and surprisal to create something *new* and in a very dense form: What is an irrelevant *outlier* in standard machine learning tasks may be *outstanding* and important in our case. As such, poetry is intrinsically hard to model for machine learning approaches which rely on the repetition with small deviation of training data.

Our immediate next steps will be to import the full corpus from Lyrikline (and more), to set up a management database for existing meta-data and to build our web-based interface. Based on this, we will start to analyze the stability of potentially significant features of prosodic style across different poems (and recordings) of one poet and next within pre-established and annotated literary styles. This step will result in the development of further higher-level features that we will annotate and use to enhance our models.

Acknowledgements The authors would like to thank Manfred Stede, Claudia Müller-Birn, Arthur Jacobs, Wolfgang Menzel and Robert Tolksdorf for fruitful discussions of the topic.

---

further indicators, for which, however, we still build our feature extraction; the example merely shows our guiding principle.

References

Steve Abney. 1991. Parsing by chunks. In R. Berwick, S. Abney, and C. Tenny, editors, *Principle-based Parsing*, pages 257–278. Kluwer.

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

Klemens Bobenhausen. 2011. The Metricalizer^2–automated metrical markup of German poetry. *Current Trends in Metrical Analysis, Bern: Peter Lang*, pages 119–131.

William Cohen. 1995. Fast effective rule induction. In *Proceedings of the 12th International Conference on Machine Learning*, pages 115–123. Morgan Kaufmann.

Gordon Burns Cooper. 1998. *Mysterious music: Rhythm and free verse*. Stanford University Press.

Rodolfo Delmonte and Anton Maria Prati. 2014. Sparsar: An expressive poetry reader. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 73–76, Gothenburg, Sweden, April. ACL.

Annie Finch. 2000. *The ghost of meter: Culture and Prosody in American Free Verse*. University of Michigan Press, Ann Arbor.

Rosemary L. Gates. 1985. The identity of American free verse: The prosodic study of Whitman's 'Lilacs'. *Language and Style*, 18(Summer):248–276.

Rosemary L. Gates. 1987. Forging an American poetry from speech rhythms: Williams after Whitman. *Poetics Today*, 8(3):503–527.

Alan Golding. 1981. Charles Olson's metrical thicket: Toward a theory of free-verse prosody. *Language and Style*, 14:64–78.

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*.

Arne Köhn, Florian Stegen, and Timo Baumann. 2016. Mining the Spoken Wikipedia for speech data and beyond. In *Proceedings of LREC 2016*.

Jana Lüdtke, Burkhard Meyer-Sickendieck, and Arthur M Jacobs. 2014. Immersing in the stillness of an early morning: Testing the mood empathy hypothesis of poetry reception. *Psychology of Aesthetics, Creativity, and the Arts*, 8(3):363.

Burkhard Meyer-Sickendiek. 2012. *Lyrisches Gespür. Vom geheimen Sensorium moderner Poesie*. Fink, Munich.

Gregor Möhler. 2001. Improvements of the PaIntE model for f0 parametrization. Technical report, Institut für Maschinelle Sprachverarbeitung (IMS), Stuttgart, September.

Charles Olson. 1966. *Selected Writings of Charles Olson*. New Directions, New York.

Marjorie Perloff. 1983. To give a design: Williams and the visualization of poetry. In Carroll F. Terrell, editor, *William Carlos Williams: Man and Poet*, page 159–186. The National Poetry Foundation, Orono, USA.

Andrew Rosenberg. 2010. AutoBI – a tool for automatic ToBI annotation. In *Proceedings of Interspeech*, pages 146–149.

Marc Schröder and Jürgen Trouvain. 2003. The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology*, 6(3):365–377, October.

Jon Silkin. 1997. *The life of metrical and free verse in twentieth-century poetry*. Macmillan, Basingstoke.

Kim Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg. 1992. ToBI: A standard for labeling english prosody. In *Second International Conference on Spoken Language Processing*, Alberta, Canada, October.

Paul Taylor. 1998. The TILT Intonation Model. In *Proceedings of the ICSLP 1998*, pages 1383–1386.

Willie Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, Evandro Gouvea, Peter Wolf, and Joe Woelfel. 2004. Sphinx-4: A flexible open source framework for speech recognition. Technical report, Sun Microsystems, Inc., Mountain View, USA.

Ian. H. Witten and Eibe Frank. 2000. *Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann, Rhodes, Greece.