10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Free Verse and Beyond: How to Classify Post-modern Spoken Poetry

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

[1]Department of Literary Studies, Free University of Berlin, Berlin, Germany
[2]Department of Informatics, Universität Hamburg, Hamburg, Germany
{hussein,bumesi}@zedat.fu-berlin.de, baumann@informatik.uni-hamburg.de

Abstract

This paper presents the classification of rhythmical patterns detected in post-modern spoken poetry by means of machine learning algorithms that use manually engineered features or automatically learnt representations. We used the world's largest corpus of spoken poetry from our partner *lyrikline*. We identified nine rhythmical patterns within a spectrum raging from a more fluent to a more disfluent poetic style. The text data analyzed by a statistical parser. Prosodic features of rhythmical patterns are identified by using the parser information. For the classification of rhythmical patterns, we used a neural networks-based approach which use text, audio, and pause information between poetic lines as features. Different combinations of features as well as the integration of feature engineering in the neural networks-based approach are tested. We compared the performance of both approaches (feature-based and neural network-based) using combinations of different features. The results show – by using the weighted average of f-measure for the evaluation – that the neural networks-based approach performed much better in classification of rhythmical patterns. The important improvement of the classification results lies in the use of the audio information. The integration of feature engineering in the neural networks-based approach yielded a very small result improvement.

Index Terms: free verse poetry, modern and postmodern poetry, rhythmical patterns

1. Introduction

Modern poetry was deeply influenced by American poets like Ezra Pound, and Pounds famous call for a new poetic rhythm, composed “in the sequence of the musical phrase, not in the sequence of a metronome” [1, p. 3]. In literary studies, such new rhythms are defined as belonging to the “prosody of free verse”, according to a definition of Charles Hartman, who claimed that “the prosody of free verse is rhythmic organization by other than numerical modes” [2, p. 13]. Whereas such numerical modes rather belong to the idea of metrical schemes such as iambic or trochaic, modern poets created new rhythms beyond these metrical schemes, no longer based on the metrical foot, but on a new idea of rhythmic fluency.

In her analysis of such a modern free verse, Marjorie Perloff draw a further line between such poetry based on modern free rhythm on the one hand, and so-called “Non-Linear Poetries”, where “the flow of the line as the individual's breath as well as of the simulation of the eye's movement from image to image, observation to observation, is inhibited by any number of ‘Stop’ signs” [3, p. 102]. In other words: The flow of reading is interrupted in a peculiar way in poems that are oriented beyond the free prose prosody. For example, by ellipses, by montages of quotations, or by a strong deviation from normal language.

Modern poetry developed a huge amount of free verse poems and a huge amount of poems overcoming the features of such free verse. The aim of this paper is to capture the whole range of such new rhythms in modern and postmodern poetry. We will investigate the prosodic features of free verse poetry and beyond by classifying a number of rhythmical patterns. We start with those patterns based on “the flow of the line” and end with those based on a “number of ‘Stop’ signs”. We took up Perloffs differentiation between free verse prosody and beyond, developing a fluency/disfluency spectrum according to this differentiation.

We illustrate this prosodic spectrum of fluency/disfluency by ranking nine different poetic styles within the free verse spectrum. (1) **Cadence**: the cadence is the most fluent one. The basic idea of the cadence is the “breath-controlled line” as an isochronous principle. Ezra Pound, who invented the idea of the cadence, was influenced by Chinese poetry, which lacks any enjambments [4, p. 358]. This explains the so-called line-sentence as the fundamental principle of the cadence. In difference to the first class, more dis-fluent poems use “weak enjambments” separating the nominal phrase and the verbal phrase of a sentence. Such “weak enjambments” can be divided furthermore according to the emphasizing of enjambments. (2) **Parlando**: the poems in the parlando style use “weak enjambments” that do not emphasize the enjambments. (3) **Variable foot**: in contrast to parlando, the poems in the variable foot class emphasize the enjambments [5]. These two classes are also rather fluent ones, compared to those poems using “strong enjambments”, because the break in the reading is not really irritating as long as it is based on regular pauses in speech. Figure 1 shows an example of the rhythmical pattern “variable foot” for the first two lines (“Während es anfängt zu schneien” \ “schaukelt das Mädchen im Hof”) (english: “While it starts snowing” \ “the girl swings in the yard”) of a poem by the German poet Harald Hartung, reading his poem “Blick in den Hof” (english: View into the yard). As becomes obvious by the intensity or power contour, the poet makes a short stop after each colon of the sentence, imitating the regular pauses in speech by taking into account the poets breathing as a rhythmical feature (pause between words “schneien” and “schaukelt”). The blanks before and after the word “anfängt” are not considered. We used a text-speech aligner for the detection of word and sentence boundaries (see section 3.1).

A more disfluent kind of poetry uses strong enjambments, which means it separates articles or adjectives from their nouns or even splits a word across lines, like in Paul Celans poems. Poems using “strong enjambments” can also be divided according to the emphasizing of enjambments. (4) **Hook style** (German: Hakenstil): the poems in the hook style use “strong enjambments” that do not emphasize the enjambments. (5) **Gestic rhythm**: in comparison to the hook style, the poems in the gestic rhythm emphasize the enjambments, which means that the author makes an irritating break after each line when reading his poem, although the sentence should continue [6]. This fifth pattern was invented by Bertolt Brecht in his later work and had a huge impact on poets from the former German Democratic Republic.

10.21437/SpeechProsody.2020-141

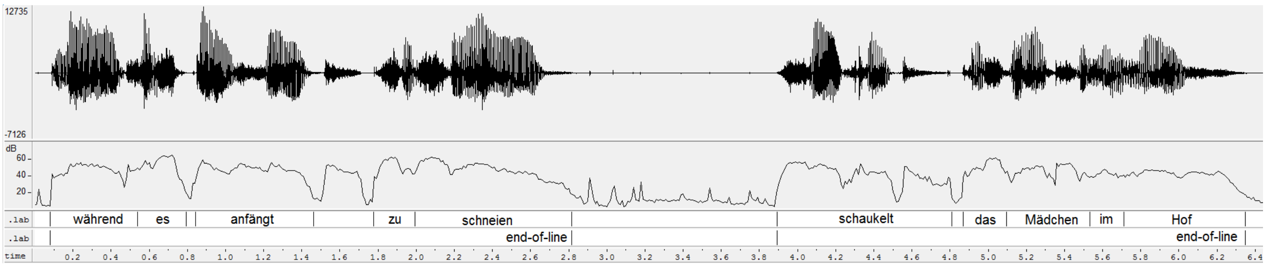

Figure 1: *Analysis of the first two lines in the poem "Blick in den Hof" (english: View into the yard) from the poet Harald Hartung as an example for the "variable foot" pattern (from top to bottom: speech signal, intensity (dB), word boundaries, line boundaries, and time).*

Moving forward towards to the radical dis-fluent pole, the next pattern is the ellipsis. (6) **Ellipsis**: is the omission of one or more grammatically necessary phrases [7]. This rhetorical figure can also affect the prosody of a poem, which has been observed in poems of Paul Celan. (7) **Permutation**: is a conversion or exchange of words or parts of sentences or a progressive combination and rearrangement of linguistic-semantic elements of a poem, a principle that was very popular in German "concrete poetry" [8]. Even more radical kinds of poetic disfluency have been developed in modern "sound poetry" by dadaistic poets like Hugo Ball and Schwitters or concrete poets like Ernst Jandl. Within the genre of sound poetry, there are two main patterns: (8) **Syllabic decomposition**, dividing the words into syllables; and (9) **Lettristic decomposition**, the last and most disfluent pattern, which divides the words into single characters, like for example in Ernst Jandl's famous poem "schtzngrmm" [9].

Using the world's largest corpus of readout poetry from our partner *lyrikline* (`www.lyrikline.org`), we identified the nine rhythmical patterns mentioned and extracted their prosodic features. We compared this manual feature-based classification with an approach based on neural networks (NNs) for the classification of these rhythmical patterns. The paper is organized as follows: An overview of the database is provided in Section 2. Section 3 describes the processing tools as well as feature engineering- (rule-based) and neural networks-based approaches. The experimental results are given in Section 4. Finally, conclusions and future works are presented in Section 5.

2. Data

We used German poems available on the website of *lyrikline* in the project *Rhythmicalizer* (`www.rhythmicalizer.net`). *Lyrikline* was initiated by the Literaturwerkstatt Berlin and houses contemporary international poetry as texts (original versions and translations) and the corresponding audio files. All the poems are read by the original authors. There are 240 German-speaking poets (Germany, Switzerland, and Austria) reading $2,657$ German poems out of a total of $1,412$ poets and $12,664$ poems on *lyrikline* (statistic from the 19th of December 2019).

An expert in literary studies (the philological scholar of our project - second author) listened to the audio recordings of 285 poems (sometimes with looking at the waveform of audio files) and classified them to one of the nine rhythmical patterns of modern and post-modern poetry (seven poems are classified into two rhythmical patterns). Table 1 show some key descriptive statistics of the poems as assigned to their classes. The total number of German poems in the table is according to the state of the 1st of April 2018.

Table 1: *Description of the data used in the experiment.*

	poems	lines	characters	audio
lyrikline: German subcorpus	2392	61849	2025484	52 hour
cadence	31	1079	34163	49 min
parlando	34	1435	44323	67 min
variable foot	34	878	23684	39 min
hook style	36	1090	33178	48 min
gestic rhythm	33	897	27741	44 min
ellipsis	56	2154	62704	104 min
permutation	30	1117	32041	46 min
syllabic decomposition	21	540	12390	26 min
lettristic decomposition	17	684	10007	31 min

3. Method

In our previous works [10][11], we classified rhythmical patterns with two separate approaches. The first one is feature-based approach. Feature engineering, for example, pause features (the average pause length at the end of each line as well as between words) and parser features (the poem's number of lines, number of lines with a finite verb, and number of lines with punctuation) were utilized. This approach used classical machine learning algorithms, for example, boosting algorithm (AdaBoostM1), Instance-Based classifier (IBk), logistic model trees (SimpleLogistic), and RandomTree from WEKA toolkit [12]. The second approach is based on neural networks (see the description in section 3.3). Using the both approaches, we classified two patterns (parlando and variable foot) in [10] and six patterns (parlando, variable foot, hook style, gestic rhythm, syllabic decomposition, and lettristic decomposition) in [11]. The neural networks-based approach yielded better results than feature-based classification approach by using the weighted f-measure for the evaluation of classification results.

In this work, we developed a feature engineering-based approach which uses different features in comparison to previous works. We integrated these features into our previous neural networks-based approach and used several feature sets in the neural networks-based approach.

3.1. Processing Tools

The following tools were utilized for the analysis of audio and text data as well as for feature extraction:

- **Text-Speech Aligner**: A forced-alignment of text and speech for poems is performed using a text-speech aligner [13] which employs a variation of the SailAlign algorithm [14] implemented via Sphinx-4 [15]. The line boundaries (the start of the first word and the end of the last word in each poetic line) are detected. 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 (the white-space is important as the text in a poem and help readers know how to read a poem out loud). The forced alignment of text and audio in spoken poetry, especially in concrete and sound poetry, is non-trivial and often individual words or lines cannot be aligned. Therefore, the automatically extracted alignment information is manually corrected by the first author more than once (rectifying alignment information and in some cases correcting of written text and audio files of poems) by listening to the audio file and looking at the waveform.

- **Parser**: We processed the text data of poems by using a statistical parser in order to extract syntactic features. The Stanford parser [16] is used to parse the written text of poems. The parser used the Stuttgart-Tübingen-TagSet (STTS) table developed at the Institute for Natural Language Processing of the University of Stuttgart [17] for the parsing of German poems. The main problems in poem parsing involve the absence of punctuation marks [7]. In addition, many poems are written with special characters: sometimes the text is written in lowercase with some words in uppercase, which makes the recognition of sentence boundaries quite difficult. Furthermore, some sentences within the poems go beyond the line boundary and run on to the next line. Such unconnected syntactic elements result from the dissolution of poetic lines, caused by so-called enjambment. No countermeasure were taken to solve these problems, since the correction of the problems is time-consuming and changes the poem form.

3.2. Feature Engineering-based Approach

We processed every poem individually, line by line, even if there are run-on lines (enjambments) within a poem, in order to extract features for the recognition of the rhythmical patterns of poems. The most important indicators for some rhythmical patterns, for example, the concrete poetry (permutation) and sound poetry (syllabic decomposition and lettristic decomposition), are the absence of a verb within a complete sentence or half-sentence and the existence of asemantic material. We used parser information that comprises abbreviations of words' Part-of-Speech to extract syntactic features of words in every poetic line.

Different features are extracted and analyzed by using the parser to parse each line in isolation. We focused on the following five inflected verbs: finite verbs (VVFIN), imperative verbs (VVIMP), auxiliary verbs (VAFIN), auxiliary imperative verbs (VAIMP), and finite modal verbs (VMFIN). We also identified the following types of nouns: normal noun (NN) and proper name (NE). We identified the punctuation marks in order to differentiate between concrete as well as sound poetry on the one hand and a rather "normal" poetry using "regular" language which is equivalent to grammatical sentences on the other hand, because complete sentences in lines can be discerned by sentence-ending punctuation (. ? ! ; :), and clauses by commas (,). Therefore, we found all the punctuation marks in every poetic line. Two types of conjunctions are distinguished: subordinate conjunction in a sentence (KOUS) and coordinating conjunction (KON). Foreign language material (FM) as well as non-words (XY) are categorized as asemantic material. The most important indicators for strong enjambments are: attributive adjective (ADJA), definite or indefinite article (ART), and attributive possessive pronoun (PPOSAT).

However, parsers cannot yet distinguish between nominative and accusative, so the most important indicator for a complete sentence is the verb. The features are recorded as follows: If the poetic line contains one or more verbs (from the five defined verbs above), a character of "y" (yes, there is a verb in the poetic line) is added to the "verb" feature in the feature vector; otherwise a character of "n" (no verb in the poetic line) is added. The same process is implemented in every poetic line for: noun, comma, sentence-ending punctuation, conjunction, asemantic material, and indicators for strong enjambment. The feature vector of the first poetic line ("Während es anfängt zu schneien") showed in Figure 1 is: ('verb':'y', 'noun':'n', 'comma':'n', 'sentence ending':'n', 'conjunction':'y', 'foreign word':'n', 'strong enjambment indicator':'n').

3.3. Neural Networks-based Approach

In this work, we used the same approach described in [18][19] for the classification of prosodic styles. The approach is based on neural networks and will be explained briefly in this section. The model must deal effectively with data sparsity, since there are a broad variety and a relatively small number of poems in the experiment. Therefore, we use as few free parameters as possible that need to be optimized during training. For this reason, we focused in the textual processing on character-by-character encoding of poetic lines (and using character embedding).

The textual information, the spoken recitation on the line level, and the information regarding pauses between lines are utilized. We extract Mel-frequency cepstral coefficients (MFCC) for every 10 msec of the audio signal as well as fundamental frequency variation (FFV) [20] vectors, which are a continuous representation of the speaker's pitch. We z-normalize all feature dimensions. In order to not overwhelm the model with acoustic sequence information, and given that relevant speech phenomena are typically much longer than 10 msec, we compute the mean and standard deviation of 10 consecutive frames for every feature. We train a neural network that learns to derive and represent features (in a multi-dimensional representation) relevant for differentiating the prosodic patterns on its own. Inspired by [21][22], we build a hierarchical attention network that encodes each line of a poem using a bidirectional recurrent neural network (RNN) based on gated recurrent unit (GRU) cells [23] and *inner attention* [24]. Pre-training with additional data from the German Text Archive [25] is implemented.

The model is not trained using an explicit notion of words. Instead, it may implicitly encode word-level information (such as Part-of-Speech) via the constituting sequences of characters. This is in line with recent work on end-to-end learning, for example, in speech recognition [26][27], which no longer explicitly models phonemes or 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 white-space,for example, indentation to create justified paragraphs—nor special characters.

We combine the line-by-line representations by another bidirectional recurrent layer 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 2. Our model is implemented in *DyNet* [28].

3.4. Experimental Setup

For the automatic classification of the nine rhythmical patterns, different feature sets from feature engineering- as well as neural networks-based approaches are extracted to build a model that

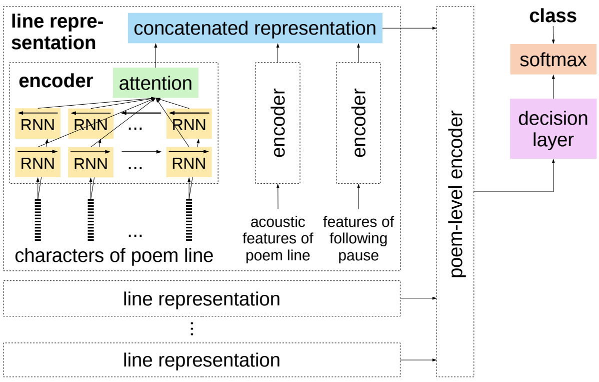

Figure 2: *Full model for poetry style detection using neural networks.*

differentiates these rhythmical patterns. The following feature sets are utilized:

- **FE**: the feature vector consists of the following features of parser information: verb, noun, comma, sentence ending, conjunction, foreign word, strong enjambment indicator (FE is for Feature Engineering).
- **NNs 1**: the feature vector consists of only text information of poems.
- **NNs 2**: only the audio information is used in the feature vector.
- **NNs 3**: the information of both text and pause (between poetic lines) is used.
- **NNs 4**: the feature vector contains text and audio information of poems.
- **NNs 5**: the audio information as well as pause between lines are used.
- **NNs 6**: the current feature vector includes the whole information of poems: text, audio, and pause.

We used the above-mentioned feature sets to classify the nine rhythmical patterns. To deal with the low amounts of data, we use 10-fold cross-validation.

4. Experimental Results

The classification results are calculated using the weighted average of f-measure. Table 2 shows the classification results of nine rhythmical patterns using FE- and NNs-based approaches. The utilization of only text information for the classification in the FE-based approach (parser information) or NNs-based approach (NNs 1) yields the worst results. The most improvement of classification results lies by using the audio information (see results of feature set (NNs 2) in comparison with results of the feature set (NNs 1)). The adding of pause information do not play an important role. The adding of parser information into the features in NNs-based approach improved slightly the results (see the results in the last column (NNs & FE)). The best results are obtained by using text, audio, and pause information as well as by adding the parser information to the text, audio, and pause information (f-measure in the feature set (NNs 6) is 0.49 and 0.50, respectively).

Table 2: *Results (weighted average of f-measure) obtained with the 10 fold cross-validation for the classification of nine rhythmical patterns by using different feature sets.*

Feature set	only NNs	NNs & FE
FE (Parser)	-	0.20
NNs 1 (Text)	0.23	0.24
NNs 2 (Audio)	0.45	0.47
NNs 3 (Text & Pause)	0.23	0.25
NNs 4 (Text & Audio)	0.48	0.48
NNs 5 (Audio & Pause)	0.47	0.50
NNs 6 (Text & Audio & Pause)	0.49	0.50

5. Conclusion and Future Works

In this paper, we have demonstrated an approach for classifying modern and postmodern poetry using rhythmic-prosodic characteristics. Taking up Marjorie Perloffs analysis of free verse poetry, we identified a number of rhythmical patterns within a spectrum that - according to Perloff - ranges from a more fluent to more disfluent “non-linear” structure. These results in a total of nine different classes, ranging from the very fluent idea of the so-called “cadence” developed in American Imagism, towards the very disfluent idea of a “lettristic decomposition” developed after 1945 by the French poet Isidore Isou and taken up by poets like Ernst Jandl. We performed forced-alignment of text and speech data of poems to detect line boundaries and processed the text data of poems by using the Stanford parser. We identified prosodic features of rhythmical patterns by using the parser information. In addition, we used a neural networks-based approach for the classification of rhythmical patterns. This approach used text and audio information of poems as well as pause information between poetic lines. We used different combinations of features in the neural networks-based approach as well as integrated feature engineering in the neural networks-based approach in order to study the effect of these different features on the classification task. We compared feature-based and neural networks-based approaches and found that the NNs-based approach performed much better in classification of rhythmical patterns. The integration of feature engineering in the neural networks-based approach yielded only a very small result improvement (for example, improvement the f-measure of 0.03 by NNs 5).

We implemented a tool which supports the human-in-the-loop approach by adding further classes to the nine classes explained above [29]. We are planning to do automatic classification on all such classes in the nearest future. The neural networks-based approach could also be used in the future to evaluate the quality of poetry translations, as a similar aspect of fluency and dis-fluency within the source (original) text and its translation, which became important since Lawrence Venuti. Venuti's critique on modern translations causing a very problematic “invisibility” of the translation itself and is based on the “fluency” of the target (translated) language. Criticizing this modern idea of a fluent target language, Venuti claims that a translation reads “fluently, when it gives the appearance that it is not translated” [30, p. 4]. The developed approach could help to identify the degree of fluency in source and target poems.

6. Acknowledgements

This work has been funded by the Volkswagen Foundation through the program ‘Mixed Methods in the humanities? Funding possibilities for the combination and the interaction of qualitative hermeneutical and digital methods’ (funding code 91926 and 93255).

7. References

[1] E. Pound and L. Pound, *Literary Essays of Ezra Pound*, T. Eliot, Ed. New York: New Directions, 1954.

[2] C. O. Hartman, *Free Verse: An Essay on Prosody*. Princeton University Press, 1980.

[3] M. Perloff, "After Free Verse: The New Non-Linear Poetries," in *Close Listening: Poetry and the Performed Word*, C. Bernstein, Ed. Oxford Scholarship, 1998.

[4] G. Steiner, *After Babel: Aspects of Language and Translation*. Oxford University Press, 1975.

[5] H. Hussein, B. Meyer-Sickendiek, and T. Baumann, "Tonality in Language: The "Generative Theory of Tonal Music" as a Framework for Prosodic Analysis of Poetry," in *Proc. of the 6th International Symposium on Tonal Aspects of Languages (TAL)*, Berlin, Germany, June 2018.

[6] ——, "Automatic Detection of Enjambment in German Readout Poetry," in *Proc. 9th International Conference on Speech Prosody 2018*, Poznań, Poland, June 2018.

[7] ——, "How to Identify Elliptical Poems Within a Digital Corpus of Auditory Poetry," in *Proceedings of the 30th Conference on Electronic Speech Signal Processing (ESSV 2019), Dresden, 6-8 March 2019*, 2019.

[8] U. Ernst, "Permutation als Prinzip in der Lyrik," *Poetica*, vol. 24, no. 3/4, pp. 225–269, 1992.

[9] B. Meyer-Sickendiek, H. Hussein, and T. Baumann, "Recognizing Modern Sound Poetry with LSTM Networks," in *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2018*, A. Berton, U. Haiber, and W. Minker, Eds. TUDpress, Dresden, 2018, pp. 192–199.

[10] T. Baumann, H. Hussein, and B. Meyer-Sickendiek, "Analysis of Rhythmic Phrasing: Feature Engineering vs. Representation Learning for Classifying Readout Poetry," in *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Association for Computational Linguistics, 2018, pp. 44–49. [Online]. Available: http://aclweb.org/anthology/W18-4505

[11] B. Meyer-Sickendiek, H. Hussein, and T. Baumann, "Analysis and Classification of Prosodic Styles in Post-modern Spoken Poetry," in *INF-DH-2018*, M. Burghardt and C. Müller-Birn, Eds. Bonn: Gesellschaft für Informatik e.V., 2018.

[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[13] T. Baumann, A. Köhn, and F. Hennig, "The spoken wikipedia corpus collection: Harvesting, alignment and an application to hyperlistening," *Language Resources and Evaluation*, 2018, special Issue representing significant contributions of LREC 2016.

[14] A. Katsamanis, M. Black, P. G. Georgiou, L. Goldstein, and S. Narayanan, "SailAlign: Robust Long Speech-Text Alignment," in *Proc. of Workshop on New Tools and Methods for Very-Large Scale Phonetics Research*, 2011.

[15] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel, "Sphinx-4: A Flexible Open Source Framework for Speech Recognition," Mountain View, CA, USA, Tech. Rep., November 2004.

[16] A. N. Rafferty and C. D. Manning, "Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines," in *Proceedings of the Workshop on Parsing German*, ser. PaGe '08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, pp. 40–46.

[17] A. Schiller, T. Teufel, C. Stöcker, and C. Thielen, "Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und größes Tagset)," Available on http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf, 1999, last accessed at 06. December 2019.

[18] T. Baumann, H. Hussein, and B. Meyer-Sickendiek, "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.

[19] ——, "Analysing the Focus of a Hierarchical Attention Network: the Importance of Enjambments When Classifying Post-modern Poetry," in *Proc. Interspeech 2018*, 2018, pp. 2162–2166.

[20] K. Laskowski, M. Heldner, and J. Edlund, "The fundamental frequency variation spectrum," in *Proceedings of FONETIK 2008*, 2008.

[21] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "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*, 2016, pp. 1480–1489.

[22] A. Tsaptsinos, "Lyrics-based music genre classification using a hierarchical attention network," in *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, 2017, pp. 694–701.

[23] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "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)*. Doha, Qatar: Association for Computational Linguistics, October 2014, pp. 1724–1734.

[24] Y. Liu, C. Sun, L. Lin, and X. Wang, "Learning Natural Language Inference using Bidirectional LSTM model and Inner-Attention," *CoRR*, vol. abs/1605.09090, 2016. [Online]. Available: http://arxiv.org/abs/1605.09090

[25] A. Geyken, S. Haaf, B. Jurish, M. Schulz, J. Steinmann, C. Thomas, and F. Wiegand, "Das deutsche textarchiv: Vom historischen korpus zum aktiven archiv," *Digitale Wissenschaft*, p. 157, 2011.

[26] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates *et al.*, "Deep speech: Scaling up end-to-end speech recognition," *arXiv preprint arXiv:1412.5567*, 2014.

[27] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International Conference on Machine Learning*, 2014, pp. 1764–1772.

[28] G. Neubig, C. Dyer, Y. Goldberg, A. Matthews, W. Ammar, A. Anastasopoulos, M. Ballesteros, D. Chiang, D. Clothiaux, T. Cohn, K. Duh, M. Faruqui, C. Gan, D. Garrette, Y. Ji, L. Kong, A. Kuncoro, G. Kumar, C. Malaviya, P. Michel, Y. Oda, M. Richardson, N. Saphra, S. Swayamdipta, and P. Yin, "Dynet: The dynamic neural network toolkit," *arXiv preprint arXiv:1701.03980*, 2017.

[29] B. Meyer-Sickendiek, H. Hussein, and T. Baumann, "Rhythmicalizer. A Digital Tool to Identify Free Verse Prosody," Available on www.geisteswissenschaften.fu-berlin.de/en/v/rhythmicalizer/Classifier/index.html, 2019, last accessed at 18. December 2019.

[30] L. Venuti, *The Translator's Invisibility*. Routledge, 1994.