

FROM FLUENCY TO DISFLUENCY: RANKING PROSODIC FEATURES OF POETRY BY USING NEURAL NETWORKS

Burkhard Meyer-Sickendiek¹, Hussein Hussein¹, Timo Baumann²

¹*Department of Literary Studies, Freie Universität Berlin, Berlin, Germany*

²*Language Technologies Institute, Carnegie Mellon University, Pittsburgh, USA
bumesi@zedat.fu-berlin.de, hussein@zedat.fu-berlin.de, tbaumann@cs.cmu.edu*

Abstract

This work offers a method for detecting the degree of fluency and disfluency used by a poet when reading his/her poems. Such a determination of (dis-)fluency plays a particularly important role in the evaluation of poetry translations. Lawrence Venuti showed that the notion of fluency became a dominating principle by which to judge English translations: a translation reads “fluently, when it gives the appearance that it is not translated” [1, p. 4]. This was meant critically, because such translations often transform an elliptic and fragmentary style within a source poem into a tangible, concrete and fluent target language. Facing current machine-translation systems, there can be no doubt that Venuti's critique is more on topic than ever: The fluency of the text in the target language became today's predominant translational ideal, due to so-called “speech disfluency removal systems” used in conversational speech translation [2]. To judge a good or bad translation thus means to estimate the degree of fluency within the original poem and its translation.

Following Venuti's critical approach, our paper will offer a new technique to estimate this degree of (dis-)fluency with regards to poetry. In a first step, we will offer a precise framework to use it for estimating a spectrum of (dis-)fluency by using two important theories for analyzing poetry: The grammatical ranking and the rhythmic phrasing. The idea of grammatical ranking was developed by Donald Wesling, whose neologism “grammetrics” is a hybridization of grammar and metrics, based on the key hypothesis that in poetry as a kind of versified language, the grammatical units (sentence, clause, group, word, morpheme) and the metrical units (syllable, foot, part-line, line, rhymated pair, stanza, whole poem) interact in a way for which Wesling finds ‘scissoring’ an apt metaphor. The grammatical ranking assumes that meter and grammar can be scissored across each other [3, p. 67]. The second important approach to detect (dis-)fluencies in poems is Richard Cureton's theory on rhythmic phrasing. Cureton divided the poetic rhythm into three components: meter, grouping and prolongation [4, p. 125]. Meter is about the perception of beats in regular patterns, grouping refers to the linguistic units gathered around a single climax or peak of prominence, quite similar to Wesling's ranking. Cureton's new idea, basically, is that of prolongation which refers to the anticipation and overshooting of a goal, the experience of anticipation and arrival. Rhythmic prolongation is a matter of connected, goal-oriented motion, based on three levels: anticipation (a), arrival (r), and extension (e) [4, p. 146]. For example: an extension occurs 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.

Using this theoretical framework, we will establish a gradual one-dimensional continuum, whose two poles are denoted by the terms “fluent” and “dis-fluent”. We illustrate this prosodic spectrum by ranking nine different poetic styles within the free verse spectrum, starting with the most fluent one, the (1 = **cadence**). The basic idea of the cadence is the “breath-controlled line”

as an isochronous principle. Ezra Pound, who invented this idea of the cadence, was influenced by Chinese poetry, which lacks any enjambments. This explains the so-called line-sentence as the fundamental principle of the cadence. In difference to this 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 into those not emphasizing the enjambments (2 = **parlando**), and those which do emphasize them (3 = **variable foot**). These two classes are also rather fluent ones, compared to those poems using “strong enjambments”. A strong enjambment separates articles or adjectives from their nouns or even splits a word across a line, like in Paul Celans poems. Poems using “strong enjambments” can also be divided into those not emphasizing the enjambments (4 = **strong enjambment**), and those emphasizing them (5 = **gestic rhythm**). Moving forward towards to the more dis-fluent pole, the next pattern is the (6 = **permutation**). A 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". The next pattern is the (7 = **ellipsis**), the omission of one or more grammatically necessary phrases. This rhetorical figure can also affect the prosody of a poem, which has been observed for example in poems of Paul Celan. 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: the (8 = **syllabic decomposition**), dividing the words into syllables; and the (9 = **letteristic decomposition**), the last and most disfluent pattern, which can be found for example in Ernst Jandl's famous poem *schtzngrmm*.

Using this spectrum, we can very accurately mark whether a translation is more fluent than the source text. Therefor we collected German poems available on the website of our partner (www.lyrikline.org). The philologist and literary scholar of the project (first author) classified 268 of a total of $\sim 2,400$ German poems into the nine prosodic classes defined above. We also collected the corresponding audio recording of each poem as spoken by the original author, yielding a total of 52 hours of audio for all German poems. We perform forced-alignment of text and speech for the poems using the text-speech aligner published by [5] which uses a variation of the SailAlign algorithm [6] implemented via Sphinx-4 [7]. This process in spoken poetry is non-trivial (in particular for decompositions in more abstract poetry). Therefore, the alignment data are corrected on the line level (start of first and end of last word for each line) as well as checked and corrected again by an expert (second author).

We present a model for the automatic classification of rhythmical patterns in the free verse poetry by using deep hierarchical attention networks. We do not use the processing on word level. Instead we used character-by-character encoding of lines in the poem and used character embeddings, sine we have a small amount of data. While processing on the word level might allow our model to build a better higher-level understanding of the poem's meaning, this semantic information would likely not help in style differentiation. In addition, word representations would not capture the usage of whitespace, for example, in indentation to create justified paragraphs or other uses, nor special characters. We use a bidirectional recurrent neural network (RNN, using gated recurrent unit (GRU) cells) which encodes the sequence of characters into a multi-dimensional representation. As for the text, we use speech line-by-line via additional encoders. We extract Mel-frequency cepstral coefficients (MFCC) for every 10 milliseconds of the audio signal as well as fundamental frequency variation (FFV) vectors, which are a continuous representation of the speaker's pitch. We z-normalize all feature dimensions. We compute the mean and standard deviation of 10 consecutive frames for every feature. To satisfy the requirement of inspectability of the decision making process, we implement a notion of *inner attention* that is to learn how to combine the sequential states of each line encodings (text, audio, and pause between lines) to a representation that is best suited towards our training

objective. We combine the line-by-line representations using a poem-level encoder which is fed to a decision layer and a final softmax to determine the poem's class. Our model is implemented in *dyNet* and *python*. Since there are a broad variety and relatively a small number of poems. We implement the pre-training with additional data from German Text Archive [8]. We used the pre-trained models in the training procedure. We first leave out the poem-level encoding and directly pass each line representation to a line-by-line decision layer. Afterwards, we replace the line-by-line decision layer with the poem-level encoder and final decision layer and train towards the per-poem decisions based on the parameters estimated before. Thus, the final model is able to steer its attention mechanism towards the important lines and can learn to sacrifice the initially trained per-line optimization for the overall per-poem optimization. Each encoder is two layers deep and has a 20-dimensional state. We train a classifier to distinguish the nine classes of poetic styles with all features (text, speech, and pause) using pre-processing and pre-training; given the little available data, we use 10-fold cross-validation (90% training and 10% test data). The best result, calculated by the average F-measure (weighted by class size), for the classification of the nine rhythmical patterns is 0.62. This indicated that it is indeed possible to check Venuti's critique of fluid translations automatically by distinguishing prosodic classes based on text, speech, and pauses using a deep neural model.

References

- [1] VENUTI, L.: *The Translator's Invisibility*. Translation Studies. London and New York: Routledge, 1995.
- [2] CHO, E., J. NIEHUES, T.-L. HA, and A. WAIBEL: *Multilingual Disfluency Removal using NMT*. In *Proceedings of the 13th International Workshop on Spoken Language Translation (IWSLT)*. Seattle, USA, 2016.
- [3] WESLING, D.: *The Scissors of Meter: Grammetrics and Reading*. University of Michigan Press, 1996.
- [4] CURETON, R.: *Rhythmic Phrasing in English Verse*. Longman, 1992.
- [5] BAUMANN, T., A. KÖHN, and F. HENNIG: *The Spoken Wikipedia Corpus Collection: Harvesting, Alignment and an Application to Hyperlistening*. *Language Resources and Evaluation*, 2018. doi:10.1007/s10579-017-9410-y.
- [6] KATSAMANIS, A., M. BLACK, P. G. GEORGIU, 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.
- [7] WALKER, W., 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*. Tech. Rep., Mountain View, CA, USA, 2004.
- [8] GEYKEN, A., 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.