Plotting Poetry: On mechanically enhanced reading, 5–7 October 2017, Basel, Switzerland

Clara Martínez Cantón, Petr Plecháč, Pablo Ruiz Fabo, Levente Seláf

The international conference Plotting Poetry: On Mechanically Enhanced Reading was organised by Anne-Sophie Bories, Hugues Marchal (both University of Basel), and Gérald Purnelle (Liège University) held in Basel, Switzerland from 5 to 7 October 2017. This conference comprised 26 presentations in English and French, delivered by scholars from eleven different countries and devoted to a wide range of projects in which poetry, poetics, and poeticity meet with computers and quantitative models.

The conference was opened by keynote speaker Franco Moretti (Stanford Literary Lab). His topic, “Totentanz. Operationalizing Aby Warburg’s Pathosformeln”, focused on visual arts, and the speaker presented their achievements in a collaborative research with Leo Impett (EPFL) aiming to build a well-defined model of the German art historian Aby Warburg’s key concept “pathos formula”. The authors proposed to model the body expression of pathos by measuring the sizes of 11 angles formed by the spine and the thighs, shins, shoulders, arms, forearms, and head of the central figures found in the Warburg’s “picture atlas” Mnemosyne – a series of thematically organised panels with nearly 1,000 photographs of paintings, sculptures, book pages, stamps, tarot cards and other types of images. Processing these multidimensional data by means of principal component analysis and cluster analysis has shown that they are able to identify not only the prototypical expression found in some motifs (“nymphs” × “non-nymphs”) but also the pathos expression of human body in general.

The second keynote of the conference, Valérie Beaudouin (Télécom ParisTech) presented the methods and results of metrical analysis of hexameter in classical French drama and 19th century French poetry. The corpus analysed consists of ca. 120,000 verses. The three possible strategies of the overviewed computer-assisted analysis are parsing the graphic chain, using machine learning techniques, or the syntactic analysis of the verses and their phonetic transcription. The software Metrometre developed by the speaker could identify metrical syllables, metrical vowels, ends of words, parts of speech and stresses, allowing the representation of the alexandrine patterns: the frequency and occurrences of the schwa, the distribution of ends of words or stresses along the verse-lines. The chronological representation of the data

https://doi.org/10.12697/smp.2017.4.2.05
showed the evolution in the use of the schwa from Corneille to Mallarmé, with a higher number of “e muet faible” in the later period, and a much more frequent use of function words at the end of the first hemistich by Rimbaud and Mallarmé than any previous poets of the studied corpus. A text mining tool permitted the clustering of the statements of the corpus dramas, and by this the distinction of different “world views”. There are clusters like “glory and honour”, “victory or loss”, “love”, etc., and their correspondence with specific rhythmical units of the hemistichs could be also represented. The clustering helped to identify not only two major topics in Racine's tragedies (love and death), but also the stylistic and grammatical patterns of their textual representations. Another result was the discovery of frequent word-plays with the rhymes of the protagonists’ names in these texts. The concluding remarks cautioned against the overestimation of the rule of digital tools in the interpretation of the texts. All hypotheses suggested by digital analysis must be confirmed by a close reading and manual verification of the data. On the other hand the computational analysis might be also a source of inspiration for poetical creation: poetic rules in form of algorithms shall be used for automatic generation of poems by computers, or they might be reused by authors in their artistic processes. Two examples were given of such an inspiration, namely, two novels written by the authors of Oulipo. Jacques Roubaud’ “Poésie”: is built up as a structure of metasonnets, where each component consists of 14 elements. The order of the chapters in Italo Calvino's If on a winter's night a traveler reproduces in a similar way the metrical structure of the endecasillabo, the most prestigious Italian verse used by Renaissance epics like Ariosto's Orlando furioso, which was an important source of inspiration for Calvino.

Muriel Louâpre (Paris Descartes University) and Hugues Marchal (University of Basel) delivered a presentation entitled “Modeling and Visualizing the Evolution of Modern French Scientific Poetry”. It took as a starting point the confrontation of the theoretical thinking of Sainte-Beuve (1835) and Patin (1848) about the death of scientific poetry with Romanticism, opposed to others like Fusil (1917), who claims that scientific poetry has never stopped flourishing, or Caro (1878), who stated that his contemporary poets were reviving it. Louâpre and Marchal’s proposal is focused in testing the validity of these perceptions and approaching the subject from different perspectives. Doing so, they presented a historical overview of the genre, which can also be useful as a case-study for understanding the dynamics of genre development. They quantified and visualized the number of scientific poetry books published per year in the 19th century, comparing two different databases. Both sources basically serve to highlight the growth of scientific poetry in the second half of the century. The authors also approached the genre
through a visualization of scientific poetry by discipline, in different periods of the nineteenth century, finding relevant and changing trends. Whereas at the beginning of the century, scientific poetry books were mostly about engineering, from 1810 to 1840 scientific poetry focuses mainly on medicine. The industrial age increased the number of poetry books devoted to engineering and earth sciences, and at the end of the century there was a clear tendency towards the poetry of biology, closely linked to that of medicine. To give a more complete approach to scientific poetry during the 19th century, the study also considered a sociological approach, quantifying and visualising the authors’ occupations in the different disciplines of scientific poetry, their geographical origin, their age, etc. Louâpre and Marchal also pointed out the possibility of focusing quantitative analysis on the formal aspects of the genre and its evolution. Overall, the talk traced a clear history of the scientific poetry genre in France in the 19th century, based on a quantitative method.

The presentation by Véronique Montémont (ATILF and ITEM labs in Nancy and Paris) was called “Machiner la poétique”, and offered a description of the poetics of the autobiography genre. Formal and content features of autobiographies were described, such as the use of first-person vs. second-person pronouns and the use of lexical items related to the concept of “authenticity”. It was discussed how these traits vary depending on factors such as subgenre within the autobiography overall genre, authors’ gender and publication period.

Clémence Jacquot spoke on the topic of the possibilities of a stylistic research of Guillaume Apollinaire’s poems assisted by textometrical tools. Apollinaire expressed his wish to simplify poetic language. This assertion was examined with the help of the TXM software developed at the ENS of Lyon. The frequency of relative pronouns was calculated, as well as the most frequent words and topics in the poet’s œuvre. The talk also discussed interpretation problems when assessing these results; the importance of human intervention in checking the results was emphasised. Finally, the talk highlighted the Hyperapollinaire project, which offers an annotated version of all poems, permitting also chronology-based queries in the corpus.

Isabelle Parkinson (Queen Mary University of London) presented on a topic entitled “A Poetics Defined in the Paratext. Significant Data in Against Expression: an Anthology of Conceptual Writing”. She focused on a network analysis approach to the content of the anthology, edited by Dworkin and Goldsmith (2011). Networks were shown depicting which authors are cited in the text introducing other authors’ sections in the anthology. The temporal distribution of cited authors (from the 1910s to the 2010s) was also analysed.
The talk of Klemens Bobenhausen (independent researcher, Freiburg im Breisgau) and Benjamin Hammerich (ETH Zürich) was entitled “Analyzing poetry. Who’s Better? Humans or Machines?” and devoted to the error rate and typology of errors found in Frank 1993, Schlawe 1972, and in the output of automatic verse processing system Metricalizer that is being developed by the speakers (http://metricalizer.de). Comparing the data on German iambic pentameter from these three sources, the speakers concluded that, although error rates of manual and automatic scansion are estimated to be more or less equal, there is an important difference in the quality of these errors. While most of them in Schlawe and Frank are produced by inertia of the annotator towards changes in the poem’s structure and/or basing the annotation on scansion of just a piece of the poem, half of the errors produced by Metricalizer were simply due to typos in the input data (the second half due to the wrong interpretation of prosodically ambiguous words). The speakers also provided a strong criticism to Frank’s book for its lack of any reference to the underlying data and bias, probably caused by a very subjective way of collecting them.

Levente Seláf’s (ELTE, Budapest) speech contrasted two methods for building poetical databases. The best results in computer-assisted analysis are obtained with full-text databases, containing already pre-treated texts, simplified from a philological point of view, belonging to “closed corpora” according to the terminology of the author: In these experiments the textual and metrical variability of the poems attested in different versions is generally disregarded, the number of poems is fixed, and the major prosodic rules are evident. Such a simplified, polished corpus can easily be transformed to TEI-XML and the analysis can sometimes obtain spectacular results. But when the poetical rules which are supposed to give coherence to a corpus are not homogeneous, and not even clear enough, the automation of the analysis is almost desperate. The talk presented the experiences of working with an “open corpus” during the building of the database Nouveau Naetebus, poetical repertory of Old French Non-Lyric Strophic Verse. Although the corpus is rather small, less than 1000 poems, some of them have over 50 manuscripts and some consist of several hundreds of strophes. This makes it very hard to propose a set of description criteria large enough to take into account all the principles behind their creation. And, of course, if there is no full text database to support such an analysis, it will be impossible to automate the comparison of the poems’ structures. Two methods were proposed to handle the difficulties due to the obscurity of the poetical rules in the case of this corpus. The first method was to adopt (and adjust where needed) the poetical terminology of the Leys d’amors, the mid-14th century poetical treatise of the Toulouse poetical school. This treatise is not only descriptive and prescriptive, but also contains some
combinatorial patterns, allowing us to describe poems that do not exactly fit the rules known from the poetical tradition. The second method proposed, which may also be valid for other “open” corpora, would be to prepare a set of *all* potential poetical constraints (on the levels of sounds, syllables, words, hemistichs, verses, strophes, and poems), whether they were ever practiced or not in poetry according to our knowledge. This second option is inspired from the poetics of the group Oulipo, and its invention of the “tableau Queneleieff”, an imitation of Mendeleev’s periodic table but for poetical rules.

Anne-Sophie Bories (Universität Basel) proposed a reading of metrical patterns in Aimé Césaire’s free verse poem *Cahier d’un retour au pays natal*. A prominent characteristic of French versification is the use of specific, unstable syllables, most of them containing the “mute e” /ə/. The paper presented Césaire’s use of this sound in the *Cahier*, its role in the phonetic system of the text, and compared it to Apollinaire’s use. Whilst most of the poem uses more of this characteristic syllables than one would expect, the data reveals peaks of increased and valleys of decreased use at and around the occurrences of a highly significant word for the author, the word “nègre”. This word has a different meaning in French – where it bears the racist connotations we know – and in Césaire’s native Martinique Creole – where it simply means “man”. The insulting meaning of the word is underlined many times by its phonetic environment, when a consonant follows the word and forces a two-syllable, ostentatiously French pronunciation. Progressively, as this meaning is replaced in the poem by that of the proud *nègritude*, the word begins to appear followed by a vowel, eventually appearing as a monosyllable, closer to the Creole word “nèg”. This paper exemplifies how text mining and data analysis can be used for stylistic studies, in this case to uncover a political use of language as Césaire overlaps several discordant voices in his poem.

Peter Verhaar (Leiden University) contrasted in his presentation, “The Heresy of Quantification”, the literary aspects studied by traditional criticism and those addressed by criticism based on distant reading. His talk proposes to develop statistical methods and visualisation techniques which can genuinely integrate in hermeneutic processes. As a case study, the poetry of Louis MacNeice and W.B. Yeats was analysed computationally, focusing on sound effects, such as rhymes (perfect, internal, slant and semi-rhymes) and alliteration. The main idea is to study within digital humanities, the stylistic devices, rather than parameters further away from traditional criticism, such as word frequency. Verhaar carried out an interesting comparison of MacNeice and Yeats, to arrive at a critical reflection on quantification: how can we quantify certain aspects of literature? Are counts always objective? As a conclusion, Verhaar pointed out that, since algorithms are created by human beings, they
carry a strong, subjective charge on the identification of textual phenomena and their analysis. His talk defended that the results of distant reading need to be well explained and interpreted.

The talk by Gérald Purnelle (University of Liège) was called “Profilage métrique: Projet de typologie automatisée en versification française moderne et contemporaine” [“Metrical profiling: Project for automated typology in modern and contemporary French versification”]. Here he outlined several dimensions to characterize 20th century poetry in French, such as prosody, metrics and rhyme. Several markers were defined for each of those dimensions, paying attention to where different characteristics stand with respect to classical French versification between the 17th and the 19th centuries: is a classical feature being respected, modified, or abandoned, in a given poem, collection or author’s corpus? For instance, one of the features proposed in the typology is the presence of hemistichs of unequal length, or the presence of lines with a ternary (rather than binary) prosodic scheme. Statistical methods to profile authors based on characteristics like the above were presented, such as correspondence analysis and hierarchical clustering. The typology presented in the talk is intended as a starting point for developing a computational system that would automate the annotation of the characteristics on which the typology relies.

David Birnbaum (University of Pittsburgh) and Elise Thorsen (Novetta) proposed in their talk entitled “Exploring Inexact Rhymes in Russian Verse” a way to automatically detect rhymes in Russian poetic texts that comprise not only perfect matches but also imperfect ones. Their algorithm decomposes the sequences of line-final sounds into distinctive features – each line is thus represented by a string of zeros and ones (bit string). In order to estimate the probability that two lines rhyme, the edit distance between their bit string representations is calculated (i.e. the number of positions where strings differ). The speakers also briefly discussed how to identify the location of stress (which is required for the decision about how many sounds to be taken into account) in the words not found in their annotated dictionary by the inference from recurring metrical patterns.

Natalie Houston (University of Massachusetts Lowell), specialist in Victorian poetry and print culture, presented a paper on historical poetics using a computational approach. Given the historical contingency of many poetic features, her work intended to test different historical models of poetry over historical poetic practice. The talk was focused on rhyme as a historical practice with changing cultural significance. Drawing on data sources like a rhyme dictionary of the 19th century, the computational analyses produced data about rhyme patterns and lists of rhyme syllables, as well as rhyme words
and percentages of perfect and near rhyme used in the 19th century. This work on rhyme represents an exploration of a formal feature departing from a historical theory, improving the understanding of the context and poetics of the time.

Another way of rhyme detection was proposed by Petr Plecháč (Czech Academy of Sciences) in a talk entitled “Collocation-driven Method of Discovering Rhymes in a Corpus of Czech, English, and French Poetic Texts”. The described method relies on the fact that any large enough corpus of rhymed poetry inevitably contains repetition of rhyming pairs. The algorithm is thus able to identify certain portion of rhymes without any knowledge on their pronunciation by means of adaptation of the usual collocation extraction technique (line-final words that co-occur in neighbouring lines more often than would be expected by chance). This sample is then used as a training set for simple machine-learning.

Roel Smeets (PhD candidate at Radboud University Nijmegen) and Lucas van der Deijl (PhD candidate at the University of Amsterdam) presented a talk entitled “Character Centrality in Bonita Avenue”. Their presentation examined the sociology of character relations in the fictional world, taking the novel Bonita Avenue (Buwalda 2010) as a case-study. The authors carried out an analysis of the character relation networks in the novel, applying what they called a computationally assisted close reading, that is, a study of a single literary work with computational methods in order to reveal quantifiable patterns and confirm or contradict what traditional criticism points out. Their work consisted in recreating the three main characters’ social networks focusing on how strongly connected these characters were to other characters in the novel, using NER for extracting characters, and Gephi to visualize it. As a result, in this particular case, they could confirm what had been said about the male main character of the novel by critics who had performed close reading on the work. The presenters also pointed out how quantifiable data allows to provide evidence on relations such as power or influence. Smeets and van der Deijl defended the use of this kind of computational approach complementing close reading, and acknowledged the difficulty of using network analysis for a comparison of multiple novels, due to their different narrative structure.

Daniele Silvi and Fabio Ciotti (both Tor Vergata University, Rome) compared in their presentation two approaches to the analysis of themes present in a corpus: human annotation using a digital platform, and text analytics. Their talk, entitled “Computer aided thematic annotation vs. topic modeling: a comparison on Italian vernacular poetry”, is based on work done within Memorata poetis, a project focused on the thematic annotation of a corpus of Latin, Greek, Italian, and Arabic epigraphic and epigrammatic texts. The
project team carried out a thematic analysis of the texts, manually annotating motifs and thematic elements, based on a predefined thesaurus. This kind of annotation corresponds to a traditional approach of thematic analysis in literary studies, consisting in a critical activity based on human hermeneutical competence, which is a qualitative method. But the question that underlies this work is if topic modelling would yield similar results and, therefore, if it can be considered a proper proxy of literary thematic analysis. With that goal, the authors compared the sets of thematic annotations produced by human annotators, and the results of a LDA topic modelling applied to a corpus of early Italian poetry. The results of topic modeling partly coincided with those of manual annotation. However, the authors warned that LDA topic modelling for poetry is still flawed, especially for short and linguistically diverse texts as the ones found in the corpus. It does not take into account specific features of the poetic form, such as the specific location of the topic in the text. In their conclusions, the authors stated that the application of topic modelling can be justified in exploratory contexts to test preliminary hypothesis. Their results open the door to further explorations, both on the methodological and critical sides.

The speech by Christian Erwich (Vrije Universiteit Amsterdam) was devoted to a computer assisted analysis of Psalm 89 of the Hebrew Bible. One of the most crucial problems in the interpretation of the psalms is the identification of the participants (God, humans, David, etc.) because of the continual shift in person, number and gender. The method of analysis consists in identifying similar shifts in the entire Bible in order to find parallels to Psalm 89, with the aim of identifying its specific genre and eventually the speaker of verses 2–5. The “longest common subsequences” were searched and identified in the biblical text in order to find grammatically parallel structures. The result of the research was the reduction of the possible speakers of the first verses of the psalm (but the identification remained unsolved), and also the determination of the precise literary genre of the poem: the linguistic analysis proves that it shares the most common features of the other psalms, and the calculation of the ratio of parallel verses shows the strongest similarity with the korahite psalms of the Bible.

Numa Vittoz (University of Zurich) questioned the possibilities of analysing musicality of free-verse poems by the example of Yves Bonnefoy’s Rive d’une autre mort. The talk presented all the difficulties that emerge in the process of identification of prosodic rules and effects in this corpus, because of the loose, unclear definition of the versification patterns in free (or semi-free) verse. As a solution the talk proposed the creation of a phonetic database of the syllables for such a corpus, enumerating also the obstacles: a software analysing the
data should take into account so many devices and particularities of the poetical text, besides its phonetic structure, all meaningful for its poetic strength and musicality, that it still seems to be too complicated to develop such a tool despite the wonderful achievements of the digital humanities.

Clara Martínez Cantón and Pablo Ruiz Fabo (both UNED University in Madrid) presented their work carried out in collaboration with Elena González-Blanco (UNED) and Thierry Poibeau (CNRS, Lattice Lab). The talk was entitled “Automatic Enjambment Detection as a New Source of Evidence in Spanish Versification”. A system to automatically detect enjambment in Spanish, as characterized by A. Quilis (1964) and K. Spang (1983) was presented. The system relies on a Natural-Language-Processing pipeline. Based on this, detection rules identify morphosyntactic contexts for enjambment. Data on the evolution of enjambment across line-positions in Spanish sonnets (15th to 19th centuries) were presented based on the results of the automatic enjambment analysis system.

Éliane Delente and Richard Renault (both University of Caen Normandy) presented the results and the limits of poetic analysis as they experienced it via the Anamètre database, created at the University of Caen. The textual corpus comprises 12,161 modern (mainly 19th century) French poems and 100 dramas (677,267 verses encoded in TEI-XML). The metrical analysis of the texts is automated on the levels of verse lines and strophic forms. The database has so far allowed obtaining metrical statistics, rhyme dictionaries, diaereses, and drawing a repertory of metres and fixed forms. Further steps of the project include the addition of stylometric and lexicometrical analysis of texts, and the application of the database for didactic purposes. The presentation also gave important examples when the automatic treatment of the corpus was not convenient, and established three categories: 1) when human interventions are necessary to obtain a correct analysis that could not be achieved with the help of the computer; 2) types of analyses that haven’t been implemented in the program but could still enrich the database (e.g. the representation of certain metrical ambiguities); and 3) phenomena that could hardly or never be treated automatically. The greatest number of human interventions was necessary in the case of the diaeresis, badly identified by the computer in case of 7,47 % of the verses. Finally, the authors mentioned further problems that the automatic analysis of French verses may face, like the treatment of enjambments, and other rare metrical phenomena.

Caroline Ardrey (University of Birmingham) presented a talk entitled “Lectures et relectures numériques de la poésie de Baudelaire à travers la déclamation orale” [“Digital readings and re-readings of Baudelaire”]. She described current results at the Baudelaire Song Project. The project is annotating audio
from poetry readings at several levels: syllable, word, line and stanza, sense units and musical phrases. The “voice visualisations” that result are helping students understand the rules of French versification. Besides, the project is building up a public archive of performance. Examples of annotated poems were presented as well as their representations on the “Visualising Voice” platform.

Burkhard Meyer-Sickendiek and Hussein Hussein (both Free University of Berlin) presented an ambitious project “Rhythmicalizer: A Digital Tool to Identify Free Verse Prosody”. Their research team aims to automatically discover and classify rhythmical patterns found in a large body of modern and postmodern poetry. The project deals with both written texts and recordings of their recitations. Furthermore it employs text-to-speech models in order to identify specificities of poetic speech. It has been shown that the combined analysis of text and audio allows identifying philologically relevant clusters such as “cadenced rhythm”, “sprung rhythm”, or “staccato rhythm”.

The presentation by Christophe Imperiali (University of Bern) was entitled “Structures du vers dit” [“Structures of spoken verse”]. He presented a model to describe the structures for a given poet’s verse at a given time. Three levels of analysis are considered, i.e. metrical, syntactic and prosodic analyses, and criteria for prominent positions are described within each level. The model takes into account both punctuation and whitespace as prosodic cues. The prosodic hierarchy is also considered as a source for identifying prominent positions. Acoustic criteria are also part of the model, such as syllable length and intensity, pitch, and pause length.

Claus-Michael Schlesinger (University of Stuttgart) presented a talk entitled “Reading Writing Machines. Text Analysis and Generative Aesthetics in Text Generators (1955–1970).” The speaker focused not only on the algorithms of the early poem generators and the texts that they have produced (e.g. Lutz 1959), but he also analysed the original “human-produced” texts that served as a basis for setting up the input vocabularies and offered a historical contextualisation of these experiments.

Antonio Rodriguez (University of Lausanne) presented “La machine lyrique ou l’incarnation du support” [“The lyrical machine or the incarnation of support”]. This was a thought-provoking talk about the role in society of poetry and of different actors related to poetry (e.g. poets themselves and academia), paying particular attention to the role of digital technology. In this respect, he introduced the concept of the “Poetry network”, which comprises poetry-related actors, such as the internet, social media, software engineering, or academic, cultural and creative institutions, as well as poetry readers and writers interacting with the actors just mentioned. He also presented the
“Printemps de la poésie” festival [“Spring of the poets”], which has been taking place since 2016 in Switzerland. This festival will involve 100 institutions in 2018 and 400 organizers.

The final speaker Christian Hänggi (University of Basel) presented in his talk entitled “The Pynchon Playlist: a Statistical Analysis” a structured overview of musical references in novels and short stories of Thomas Pynchon. The speaker collected nearly 1,000 such references (including very cryptic ones) and offered a detailed analysis of the musical genres and composers with which they are being associated. Furthermore the speaker discussed the density of musical references in Pynchon’s specific works and its correlation with the period in which the plot takes place.¹

¹ Clara Martínez and Pablo Ruiz were supported by the POSTDATA Project (Starting Grant ERC-2015-STG-679528, PI Elena González-Blanco). Petr Plecháč was supported by the Czech Science Foundation, project GA17-01723S (Stylometric Analysis of Poetic Texts). All the authors contributed equally to this work.
References


