

Plotting Poetry 3. Conference report

David J. Birnbaum, Anne-Sophie Bories,
Thomas N. Haider, Mari Sarv*

Plotting Poetry (and Poetics) 3 / *Machiner la poésie (et la poétique) 3*, the third convocation of an international group of scholars who share an interest in the machine-assisted exploration of poetry and poetics, met on 26–27 September 2019 in Nancy at the ATILF (Analyse et Traitement Informatique de la Langue Française) Laboratory, part of the Université de Lorraine and of the CNRS (Centre National de la Recherche Scientifique). The Nancy meeting, which follows earlier ones in Basel (2017) and Berlin (2018), was organized by *Anne-Sophie Bories* (University of Basel) and *Véronique Montémont* (Université de Lorraine) and included twenty presentations in English and in French. The program and other materials from all three meetings to date are available on the Plotting Poetry website at <https://plottingpoetry.wordpress.com/>.

The first presentation of the conference, *Ophir Münz-Manor's* “Hebrew liturgical poetry from late antiquity: preliminary computational explorations”, reported on the author’s experience of revisiting with digital tools (the CATMA toolkit, which formed the focus of Jan-Christoph Meister’s keynote address the following day) the quantitative research he had undertaken manually some fifteen years earlier in the context of his PhD dissertation. Münz-Manor’s object of study was a corpus of some 220 texts (about 50,000 words) of Hebrew liturgical poetry (*piyyut*), and his focus was on the use of figurative language (especially metaphor, but also simile, synecdoche, personification, and metonymy) in those texts, which turned out to be surprisingly marginal (only slightly more than half as dense as in *Psalms*). Reanalyzing his original dissertation data with CATMA yielded some results that mirrored those obtained with manual methods, others that refuted earlier conclusions, and still others that were new and that would not have been available without the use of computational tools and methods.

* Authors’ addresses: David J. Birnbaum, University of Pittsburgh, 1228 Cathedral of Learning, Pittsburgh, PA 15260, USA, e-mail: djbpitt@pitt.edu; Anne-Sophie Bories, University of Basel, Maiengasse 51, 4056 Basel, Schweiz, e-mail: a.bories@unibas.ch; Thomas N. Haider, Max Planck Institute for Empirical Aesthetics, Frankfurt and University of Stuttgart, Grüneburgweg 14, 60322 Frankfurt am Main; e-mail: thomas.haider@ae.mpg.de; Mari Sarv, Estonian Literary Museum, Vanemuise 42, Tartu, 51003, e-mail: mari@haldjas.folklore.ee.

Helena Bermúdez Sabel and *Pablo Ruiz Fabo*, in their “DISCOVering Spanish sonnets: a close/distant reading experience” (coauthored with *Clara Martínez Cantón*, who was not able to be present in person), introduced the Diachronic Spanish Sonnet CORpus (DISCO). This corpus, which at the time of the conference included more than 4,000 sonnets spanning five centuries and drawn from canonical and non-canonical European and Latin American writers, offers, in conformity with FAIR Data Principles, open access to full TEI XML, plain text, and RDFa information through resources hosted on GitHub and Zenodo. Mediated access to information in the corpus is also provided through the DISCOVer front end, a web interface that offers predictive faceted searching and, in a *digital workstation* format, interactive textual and graphic reports about formal poetic features at user-controlled levels of granularity. DISCOVer is designed to support navigation between poem and corpus views and between close and distant reading, which the presenters illustrated by moving from an individual poem to a close view of rhyme pairs, then to specific rhymes that occur elsewhere in the corpus, then to subsetting those results according to metadata, and then, completing the circle, to other poems that contain the new rhyming words. Beyond the study of Spanish sonnets, and beyond the study of poetry more broadly, the open access DISCO corpus and the feature-rich, user-friendly DISCOVer front end exemplify an effective and economical way to repurpose the same data by offering alternative modes of access. On the one hand, the project meets the needs of technologically adept humanities researchers by offering access to raw data that is available for reuse. On the other hand, DISCOVer provides exploratory tools that enable researchers without coding expertise nonetheless to navigate structural, linguistic, and metadata features of the corpus in sophisticated interactive ways.

Mari Sarv’s “Exploring variation in Estonian folk songs: language, meter, topics, genres” reported about the analytical exploration of a corpus of approximately 100,000 Estonian *runosongs*, a formally distinctive type of Finnic folk poetry characterized by trochaic meter, alliteration and parallelism, a stichic form without rhymes and couplets, and common motifs, themes, and poetic formulae. While unified by these formal and thematic consistencies, the corpus exhibits variation in meter, word forms, and content in ways that are sensitive to, although not fully coextensive with, dialect areas (especially in the north). Archaisms and dialect features make it impossible to parse the texts using tools developed for and trained on the standard language, but Sarv’s presentation demonstrated how the analysis of poetic features (especially meter and alliteration) and the use of unlemmatized word forms (in stylometry and topic modeling) have nonetheless already produced new knowledge about patterns of variation in the Estonian folk song tradition. One outcome of this

analysis has been the successful use of corpus data (specifically, most frequent words) to document regional differences, e.g., a preference for communicative terms (*hello, thank you*) on the islands, a preference for terms for relatives (in wedding songs) in the west, etc. Quantitative analysis also reveals regional differences in vocabulary within the same semantic or thematic sphere, so that, for example, among kinship terms father, mother predominate in the north, while brother, boy, girl are more common in the south. Beyond what these studies reveal about distributional patterns in the Estonian folk tradition, though, they demonstrate that quantitative methods can help identify meaningful formal and thematic patterns even when applied to dialectally varied data that is not amenable to normalization with tools developed for standard-language texts.

Éliane Delente's "Le traitement automatique du rythme régulier. Un cas particulier: l'enjambement" ("Automated processing of regular rhythm, and the special case of enjambment") opened by asking "To what extent is the automated processing of enjambment in French versified discourse possible?" Delente grounded her engagement with this question by first explicating how a poet's use of enjambment is influenced by a diversity of features, some metrical (e.g., verse length and structure, stanza type) and others historical (e.g., time period, author, reader). Furthermore, although enjambment is traditionally understood as overlap between formal metrical properties (e.g., lines) and formal syntactic ones (e.g., sentences, phrases, or clauses), it also has semantic, pragmatic, and discourse aspects. On the discourse level, for example, listeners cannot process an entire poem, or even stanza, at once, and instead consume it in perceived *metrical units*, which are governed by expectations based on individual subjective knowledge, experience, and sophistication, on the one hand, and, on the other, also by formal features that might be amenable to automated detection. Delente categorizes metrical units as *complete* vs. *incomplete* and as *effective* vs. *retrospective*, which provides a formal apparatus to describe, for example, situations where a first line appears metrically complete because it is syntactically complete... until its continuation in the next line causes a reassessment (retrospective) that reveals it to be enjambed (much as the ellipsis does in this sentence!). While Delente concludes that enjambment is too varied and too complex for its detection to be fully automated, she proposes that corpora built from the hemistich up with information about relationships between adjacent metrical units would provide a more sophisticated resource for training and analysis than the simple dichotomy of end-stopped and run-on lines that has traditionally characterized the discussion of enjambment.

In their “Rules-based and machine-learning approaches to identifying Russian rhyme”, *David J. Birnbaum* and *Elise Thorsen* reported on recent developments in their ongoing project of identifying rhyme, with particular attention to inexact rhyme, in Russian verse. Because one goal of the project is to distinguish where similar, but not identical, sounds function as a rhyme and where the same sounds do not, the project analyzes rhyme in the context of an individual poem, without attempting to construct a global list of rhyme pairs or sets across the language. The conference presentation demonstrated the use of unsupervised learning (hierarchical agglomerative clustering) to identify the line endings that were closest phonetically within a poem, decomposing those line endings into segments and then phonetic distinctive features in order to explore and discuss rhyme behavior at a more granular level than words. While last year’s presentation of an earlier stage of this research at Plotting Poetry 2018 (Berlin) had decomposed the line into syllables, and the syllables into onset, nucleus, and coda, this year’s presentation offered evidence that decomposing syllable structure in this way may complicate the identification of rhyme, and a simpler decomposition into vowels and consonant clusters (irrespective of syllable boundaries and onset/coda distinctions) is more effective for recognizing heterosyllabic consonant correspondences that nonetheless contribute to rhyme.

Petr Plecháč’s “Yet another fresh confirmation of Mr. Spedding’s division of the play of *Henry VIII*” introduced a new way of assessing the attribution of individual scenes (or parts of scenes) of Shakespeare’s *Henry VIII* to Shakespeare or to John Fletcher. That Fletcher participated in the authorship of *Henry VIII*, first suggested by James Spedding in 1850, is by now widely accepted in Shakespeare studies. The hypothesis has been tested repeatedly with stylometric methods using a variety of features, to which Plecháč contributed new evidence in the form of the distribution of stresses within each line. Plecháč applies a support vector machine as a classifier, rhythmic types of lines (identified automatically using methods developed by Ryan Heuser for the Stanford University Trans-Historical Poetry Project), and other plays of Fletcher and Shakespeare as a training corpus, and he uses a *rolling delta* technique (developed by Maciej Eder in 2015) to explore authorship without dependency on formal scene divisions. Plecháč’s analysis produced almost perfect identification for works known to be by Shakespeare or Fletcher, and the results of classifying portions of *Henry VIII* by rhythmic line type, by 500 most frequent words, and by a combination of the two sets of mutually independent features all conform closely to Spedding’s 1850 attributions.

The first keynote of the conference, by *Anne Bandy-Scubbi*, was entitled “Zooming in, zooming out: 30 years of corpus stylistics bricolage”. The title’s

code-switching to the French bricolage (do-it-yourself and makeshift) aptly illustrates the author's tool-switching habits and her goal-driven, undogmatic, and versatile application of a wide range of computation methods. Reviewing her own history of corpus studies of eighteenth-century English literature, Bandry-Scubbi recounted how, beginning with magnetic tape, she experimented with many successive methods and perspectives (frequency counts, concordances, hyperbase, TXM, and others), and how she reused data from other teams, explored traces of authorial gender or writing strategies, and tried to pinpoint the representative and distinctive features of a canon. Drawing on and integrating traditional and novel methods, close and distant reading, and famous works and less well-known ones, Bandry-Scubbi retold a career of actively addressing the open-ended question of the hermeneutic benefit of computational and statistical approaches to literary phenomena.

The starting point for *Chris Mustazza's* "In search of the sermonic: hearing sonic genre in poetry recordings" was James Weldon Johnson's 1927 *God's trombones*, a collection of sermon-poems modeled on the speech sounds of African-American preachers of the early twentieth century, recordings of which, in Johnson's own voice, are available in the PennSound archive. Mustazza applied quantitative methods to model the pitch, tempo, and other sonic properties of these "sermonic poems" alongside those of actual sermons, with the goal of exploring whether it is possible to identify how specific acoustic features contribute to the sermonic quality of the poetry. This, in turn, invites broader inquiries about the hermeneutic use of sonic features for corpus exploration, e.g., whether it is possible to use digital methods to distinguish modernist poetry modeled not only on sermons, but also on political speeches or comedy acts. Ultimately, the association of formal acoustic properties with poem types raises the question of whether it is possible to employ digital methods to identify sonic genres and classify audio recordings automatically.

Valérie Beaudouin's "Rap et métrique" ("Rap and meter") begins by observing that since its advent in African American musical culture of the 1970s, rap has undergone transformations and adaptations as it has spread around the globe. Much as the diffusion of the sonnet from its origin in southern Italy across Europe and beyond introduced local accommodations to the particularities of new host languages, rap in different territories exhibits differences in text, in music, and in theme. These differences invite us to ask whether it is possible to model varieties of rap – with particular attention to French and francophone rap – as a confluence of speech flow and musical rhythm.

Beaudouin's research repurposes a suite of tools first developed to study the meter and rhythm of seventeenth-century French verse (Corneille, Racine,

Molière), specifically *Metrometer* (which converts orthography to phonetic representation, with stress, word boundaries, syllabification, and part-of-speech tagging) and *Rimarium* (which analyzes rhyme using graph methods, clustered by final metrical vowel, post-tonic consonants, and grammatical endings). The goal in introducing these formal annotations is to look for relationships between formal poetic features and semantic content, e.g., do we find different rhythmic structures in clusters around such different semantic environments as “love” and “death”? But lines in rap are not like Alexandrine verses (line length in rap is highly variable and there is no regular internal caesura), although they also are not free verse (there is no systematic avoidance of rhyme and caesura); end rhyme is not systematic, but internal rhyme is rich. Beaudouin concludes by suggesting that a rap corpus that is suitable for data mining should include both metrical features (placement of rhyming syllables, placement of accented syllables, degree of correspondence between syntactic units and measures, number of syllables per beat) and performative features (legato vs. staccato, degree of articulation of consonants, extent to which the onset of any syllable is earlier or later than the beat). Insofar as, in rap, an increase in tempo could accommodate an increase in the number of syllables within the same metrical space, when trying to make sense of rap it is important to start not from properties of the text, as one might with classical French verse, but from the meter and the general musical structure, including performative features.

In his “Hayford’s duplicates: operationalizing a literary theory of Herman Melville’s *Moby-Dick*”, Jonathan Armoza employed digital methods to interrogate a hypothesis by Harrison Hayford, an editor of Melville’s writings, who in his “Unnecessary duplicates” proposed that *Moby-Dick* was assembled (“cobbled together”) from the three primary draft stages, none of which survives in manuscript form. Hayford’s theory arose from a quest to explain duplicate and vestigial settings, events and characters, and his argumentation relies on stylistic evidence, inductive reasoning, and Melville’s personal letters. The three stages Hayford proposes are: 1) a sea narrative, without Ahab, Queequeg, or Bulkington, where “savages” are background, and not individuated; 2) Bulkington as teacher, comrade, and also truth-seeker (a stage for which little textual evidence remains); and 3) Melville splits Bulkington into Queequeg (mentor and friend) and Ahab (truth-seeker). Once Queequeg, a “savage”, has become a central character in stage 3, Melville injects phrases like “my particular friend Queequeg” and “my dear comrade and twin-brother”, etc. Hayford sees these phrases, which follow a part-of speech pattern of possessive + adj(s) + noun(s) or proper-noun, as clumsy insertions, a hypothesis

that might be tested by examining the distribution of the pattern across the text of the novel.

Armoza examines Hayford's theory through a prism of Nonnegative Matrix Factorization (NMF), a filtering method that has been applied productively to extracting latent information in the analysis of sparse data in non-literary domains. In the case of *Moby-Dick*, Armoza's analysis concentrates on latent part-of-speech patterns in the novel. For example, if we examine sentences where Queequeg is mentioned explicitly as comrade, harpooner, or both, NMF supports the hypothesis that there was a change in part-of-speech profiles between stage 1 and stage 3, especially with respect to pronouns. *Moby-Dick* is not, to be sure, written in verse, but to the extent that Armoza's report confirms the ability of NMF to identify latent linguistic patterns in the novel, it invites us next to explore whether the method can also be applied productively to the analysis of linguistic patterns in verse texts.

Steffen Eger and Thomas Haider reported on projects related to "Controlling style in machine-generated poetry", co-authored with Jörg Wöckener, Tristan Miller, and Luise Borek, who were unable to be present in person. More specifically, they explored the poetry generation capabilities of recurrent neural network language models (RNNLM) for German and English verse. In one investigation they tuned a GPT-2 transformer architecture on 74,000 German poems and generated metrical lines (iambic pentameter) with a generate-and-filter approach. This approach yields coherent prose and metrical lines, constrained only by mistakes of the metrical model, although the generate-and-filter approach also has several shortcomings. They also reported on experiments in conditioning a RNN with sentiment polarity lexica, and on alliteration and rhyme, to produce a smarter model, instead of having to filter for features or alter the model architecture. The results were encouraging for sentiment, but the system failed to learn phonetic features, for reasons that the authors are inclined to attribute to the volume of data and the difficulty of engineering appropriate models. Finally, the authors reported on their current efforts to annotate poetry with fine-grained aesthetic emotions, with the eventual goal of conditioning poetry generation on this annotation.

In his "Poetic program, typographic system", Rémi Forte introduced a "Call for applications" poster he had produced for the Atelier national de recherche typographique, which he employed as the framework for a poetological account of the use of typography to support the creative organization of information. The methods Forte demonstrated were grounded in Kenneth Goldsmith's "Uncreative writing", and specifically in the observation that "we are faced with an unprecedented amount of available text, the problem is not needing to write more of it; instead we must learn to negotiate the vast

quantity that exists”. At the intersection of poetic writing and graphic design, the poster’s excessively long text and its dense, monotonous layout are structured around the text of the actual call for applications, interrupted by the inclusion of text fragments found online through searching for the Call’s keywords. The poster was then typeset according to typographic methods based on Karl Gerstner’s *Designing programmes* (1964). This presentation explored issues of subjectivity in constrained productions, efficiently engaging in dialog with issues of data availability, personal interpretation, and systematic protocols in the other presentations.

In his “Ngram-driven word-level recombination: exploring a search space of metrically valid verse”, *Pablo Gervas* reflected on the importance of meaning, content, and thematic cohesion in computer-generated poetry. The method he presented departs from traditional approaches to the algorithmic construction of verse, which have often relied on the articulation and recombination of smaller existing poetry units into new verse, often producing pleasing results while failing to achieve a voice distinct from the originals used. Gervas’s innovative approach combines the rule-based analysis of possible word combinations with a word-chaining approach based on an n-gram model trained from a corpus. Once such mechanisms are in place, the procedure can be extended beyond the recombination of poems into other poems, involving, for example, also the recombination of texts of any kind into poems. This paper poses the crucial question of the uniqueness of a poet’s voice, and of the complex contradiction inherent in the notion of a Turing test for poetry generators, since being indistinguishable from other poems is precisely what a human poet might seek to avoid.

Natalie Houston’s “Distant reading the semantic field of English poetry” explored the horizon of expectation described by Hans Robert Jauss in “Literary history as a challenge to literary theory” – the framework of ideas, assumptions, and culturally encoded meaning that readers bring to a work of literature at a given historical moment. Rhyme dictionaries describe as well as prescribe rhyming practices, dictating what is perfect and what allowable rhyme, and Houston explored this topic through the hundred most common rhyme words in British nineteenth-century poetry, some of which functioned mainly as rhyme words and are not used elsewhere in the poems. The horizon of expectation that nineteenth-century English readers brought to their reading of poetry was necessarily different from that which readers today bring to the same texts. Conventional rhyme patterns, rhyme words, and rhyme sounds of English poetry constrained the words used in poetry and thus shaped its semantic field. Although a complete Victorian reading experience cannot be reconstituted, this historically inflected method for a distant reading of the

structures of rhyme, repetition, and enjambment in a corpus of 80,000 English poems (1835–1899) operationalizes nineteenth-century poetic theory and criticism. Such a computational distant reading of a large historical corpus sheds light on this semantic field at a scale impossible with human reading alone, enriching historical reading practices to better evaluate which nineteenth-century poems are truly typical and which are truly unique.

Benjamin Krautter and *Janis Pagel* presented a report on joint work with *Nils Reiter* and *Markus Willand* on “Identifying character types in German drama”, which sought to determine automatically the role type (e.g., tender father, comical old man, etc.) of 262 characters drawn from 46 plays in the German portion of the DraCor Corpus. The identification begins by using information drawn from second sources to annotate the characters. The authors then automatically classify these roles with a gradient boosting machine, based on metrics that are commonly used to determine character centrality, such as textual statistics (token/utterance counts, type/token ratio), network metrics (degree centrality), as well as topic models. While the method cannot accurately classify all types with confidence, certain type pairs, such as aristocrat ~ servant or father ~ daughter are more easily distinguished.

Jan-Christoph Meister, in his keynote lecture “From qualitative to quantitative analysis of poetry and back”, engaged with the trope of the methodological divide that is often assumed to separate qualitative and quantitative methodology. Starting from an observation that poetry enables new ways of seeing the familiar through deviations from ordinary language (cf. Jakobson’s poetic function), Meister set out to explore whether computer-assisted access to poetry has any meaningful connection with the human experience of poetry.

Based on Ast (1808) and Dilthey (1874), Meister introduced the methodological principles of the hermeneutic circle, which is iterative, rather than linear, and grounded in intersubjectivity (no single authoritative interpretation) and reflexivity (awareness of subjectivity, historical context, inexplicit contingent factors, etc.). Where traditional hermeneutics focuses primarily on close reading, new methods in text mining and natural language processing focus on distant reading that trades hermeneutic interpretation for statistical generalizability, even at the expense of historicity, genre-specificity, and other contextual detail. The context nonetheless remains hermeneutic because specificity may be no less important than generalizability for the humanities, as Meister illustrated through a historical misunderstanding of the biological origin (in Goethe) of Vladimir Propp’s use of the term *morphology* to characterize his structuralist folk tale methodology.

One grand challenge for digital humanities, then, is modeling the phenomenological-empirical continuum of understanding. To explore the continuum

between qualitative and quantitative methodology, Meister proposed manual annotation that is modeled around the hermeneutic cycle and that promotes ambiguity. As a way of working toward this type of approach he introduced the annotation platform CATMA (<https://www.catma.de>), which is designed to enable the hermeneutic (undogmatic – thus the zoologically punning name) within a digital, taxonomy-based approach to text and corpora.

The solution for the stated methodological problems then becomes the parameterization of context, that is, to make explicit the conceptual framework via meta-annotation, so that subjective context serves as a fruitful hermeneutic parameter. Insofar as we can model subjective interpretive text annotation algorithmically, we can begin to bridge the gap between qualitative and quantitative methods of digital text and corpus analysis. This perspective, then, suggests a strategy of moving between quantitative and qualitative analysis via annotation schemes.

Guillermo González Pascual's presentation on “Exploring stylistic consistency in Antonio Machado’s early and forgotten poems” examined the oeuvre of a poet associated with modernism, symbolism, and romanticism. The principal goal of the study was to ascertain why some poems from Machado’s 1903 book “Soledades” were omitted from his collected poems of 1907, whether they form a stylistically consistent group, and how they differ from the rest of his work. After starting with an overview using Voyant on the early texts of Machado that were included in and omitted from his collected poems, González applied Maciej Eder’s *rolling delta* method, producing bar-chart visualizations based on features that included word count, part-of-speech distribution, and topic modeling. These methods of distant reading revealed that the main topics in both books were largely the same, but there were slight differences in the use of some lexical elements characteristic of the literary movement of Modernismo, such as “abril”, “noche”, “luna”, and “sueño”. The stylometric analysis of the poems showed that, indeed, a part of the poems left out from the collected volume of 1907 were stylometrically consistent. The analysis also showed that one poem that should have been left out as characteristic of Modernismo was nonetheless included in the collected poems.

Clémence Jacquot's “Enjeux de l’annotation stylistique des motifs romanesques pour interroger la poétique des textes romanesque” (“Stylistic annotation of fiction traits to explore the poetics of fiction texts”) presented a classification of close subgenres of novels through the analysis of motifs. This report was part of the PhraséoRom ANR-DFG project, exploring a vast digitized corpus of novels of the second half of the twentieth century in English, French, and German, and bringing together specialists in linguistics, computer science, and stylistics. This annotation endeavor, through a statistical

analysis of fiction genres, makes it possible to identify and define phraseological and discursive units, or *features*, in order to describe and analyze the literary language of contemporary novels. The stylistic annotation of features is a delicate undertaking, as it challenges the traditional units and relevance levels of stylistic analysis: it tends to describe and, ultimately, to offer a prototype for modeling discursive functions specific to each recurring lexical-syntactic construction, statistically extracted from the corpus. The linguistic characterization of fiction is based on distinguishing para-literary sub-genres: crime fiction, historical, sentimental, sci-fi and fantasy, and a sub-corpus of “general” literature, defined by editorial criteria. The presentation argued that literary language is characterized by the statistically significant overrepresentation of lexemes (keywords), co-occurrences, and phraseologisms, and it focused first on presenting methods for the stylistic annotation of traits, and then on exploring the theoretical and epistemological negotiations at stake when questioning the utility of such annotation for the computer-assisted reading of fiction.

The presentation on “Poeticisms and common poetic discourse in the digital ‘Russian live stylistic dictionary’”, by *Georgy Vekshin* and *Marina Lemesheva* (coauthored with *Egor Maksimov*, who was not able to be present), identified two main communicative intentions: to *seem* and to *be*, reflected in different aspects of linguistic expression. The presenters demonstrated a web service that identifies stylistic connotations of words and phrases in their multidimensionality and historical instability, with particular attention to poeticisms as features of poetic discourse. Words are categorized for features like colloquial, administrative, ideological, scientific, poetic, and religious properties, which are thought to be language-independent and universal. For example, the Russian word *vdokhnovenie* ‘inspiration’ (etymologically a Church Slavonicism) is recognized as neutral, while its cognate and orthographically almost identical *vdokhnoven’e* ‘illumination’ is poetic. The presentation also explored the segmentation of communicative space (sociocultural framing) and basic types of pragmatics and communicative codes, observing, for example, that Russian tends to isolate *state* and *power* as values and codes in language. In the Live stylistic dictionary, pragmatics supplies strategies, operations, and devices for text, and a combination of multiple topics and contexts can contribute to a search query (called *styleset*), which returns output according to eighteen primitives, passed through machine learning to yield stylistically homogeneous texts as output.

As a whole, the conference discussed the three major literary macro-genres: fiction (Armoza on *Moby Dick*), drama (Krautter et al. on DraCor, Plecháč on Shakespeare), and, of course, poetry. The presentations about poetry ranged

over a variety of formal features, including rhyme (Birnbaum and Thorsen, Houston), metrical form (Beaudouin on the Alexandrine), and enjambement (Bermúdez and Ruiz, Delente). The approaches to these texts, in turn, ranged over topics including authorship analysis (Plecháč on Shakespeare), analysis of variation (Sarv on Finnic folk texts), and automated poetry generation (Gervas, Eger and Haider). And while most presentations, as has long been common in computational textual scholarship, operated with written texts, Mustazza's report on sermonic language stood out by virtue of its willingness to engage with the less well-developed but no less promising domain of audio analysis of poetic diction.