Studia Metrica et Poetica 12.2, 2025, 118 –131

Plotting Poetry 8: Skeletons in the Closet Conference Report

Neža Kočnik, Anna Mędrzecka-Stefańska, Albin Thörn Cleland*

The eighth Plotting Poetry conference, *Skeletons in the Closet*, took place in Prague, Czech Republic, from 16 to 18 June 2025. It focused on everything in poetry that remains hidden or unsettling: features that stand out, resist classification, or do not fit dominant theoretical and historical narratives. Alongside these thematic contributions, the programme also included papers that more broadly explored poetry and poeticity through quantitative and computational methods.

Emily Hedvig Olsson, *Anja Hendrikse Liu*, *Sarah Immel* and *Bríd-Áine Parnell* (all of the University of Edinburgh) opened the conference with "'Words Unformed and Unly': Computational emulations and creative explorations of Cummings' deviant morphology", presenting the results of their forays into experimental stylistics. Using simple word banks and algorithms for word formation as well as a fine-tuned small transformer (EleutherAI GPT-Neo-125m), they investigated whether the non-standard morphology and neologistic style of E. E. Cummings is internally consistent enough to be quantitatively and qualitatively replicable after training on a dataset of Cummings' complete poetic œuvre. The authors read and gauged the different models' generated poems according to both the quantifiable stylometric features associated with Cummings' poetry in the literature and what they call the poems' intuitive "Cummingsness", by which they mean the interplay of form with authorial intention and meaning. While some of the generated poems captured the desired quantifiable features, the "Cummingsness" was found strikingly lacking across the board. Turning the spotlight on themselves as readers and ultimate arbiters of meaning, they described how they still deeply engaged with, enjoyed, and felt both moved and provoked by the ersatz

* Authors' addresses: Neža Kočnik, Department of Slavic Languages and Literatures, Faculty of Arts, University of Maribor, Koroška 160, 3000 Maribor, Slovenia, email: neza.kocnik2@um.si; Anna Mędrzecka-Stefańska, The Institute of Literary Research of the Polish Academy of Sciences (Instytut Badań Literackich PAN), ul. Nowy Świat 72, 00-330, Warszawa, Poland, email: anna.medrzecka@ibl.waw.pl; Albin Thörn Cleland, The Swedish graduate school of digital philology, Lunds universitet, Box 201, 221 00; email: albin.thorn_cleland@klass.lu.se.

https://doi.org/10.12697/smp.2025.12.2.07

verses, even though these failed to give any impression of authenticity, thus problematising the very notions of "author" and "intention" in the context of machinic art.

The second talk, by *Ranveig Kvinnsland* (University of Oslo) and *Ingerid Løyning Dale* (The National Library of Norway), was titled "Searching for the obvious: mis-recognising repetition patterns in Norwegian poetry". They reminded us that while claims that the 1890s were a period of "lyrical renewal" in Norwegian poetry abound, these claims are in fact all based on close reading of mainly five canonical poets. To investigate the decade's production on a larger scale, they selected more than two thousand poems by forty-six authors and set out to algorithmically quantify the use of three types of repetition: anaphora, alliteration, and rhyme. The existence of phonemic transcription software for Norwegian enabled the circumvention of orthographic variation, which is important for correct rhyme identification. Some rhyme patterns, such as abac, were much more common in transcription than in orthography. Alliteration showed a preference for *s*, even when excluding from the corpus the extreme poem *Sirius som seér* (Gunhild Wexelsen, 1891), in which all words begin with *s*. The quantification of anaphora was complicated by the fact that its import seemed to depend on context – a subset of poems had every single line beginning in the same way, a practice which clearly does not bestow extra emphasis on any particular line, in the way the stylistic figure normally tends to do. Overall, Kvinnsland and Dale underscored that the statistical significance of their results remains inconclusive until a manually annotated gold-standard corpus of poems can be used as a baseline.

In her presentation "Froward correlations", *Yelena Sesselja Helgadóttir* (Árni Magnússon Institute for Icelandic Studies) discussed a long-standing experiment that investigated potential correlations between the internal structure of post-medieval Icelandic þulur (folk poems) and various external parameters such as the informant's gender, social status, time of recording, and place of origin. Drawing inspiration from phylogenetic approaches in stemmatology, the study encoded both the internal textual features and the metadata of 74 þulur as sequences using digits and placeholder symbols (e.g. "?" for unknown values). For each parameter, a matrix of Hamming distances was generated and then compared to the matrix of internal structural distances to calculate a correlation coefficient. Despite strong expectations based on earlier research – particularly regarding regional and temporal variation in þulur – the results were unexpectedly weak. The highest correlation was found with informant gender (0.106), while the lowest was with place of origin (0.009), the parameter where the strongest correlation had been expected. Even when the dataset was expanded and revisited ten years later with new digital

tools, including AI-assisted analysis, the overall results remained consistent. Helgadóttir reflected on several possible reasons for these surprising findings: the chosen structural descriptors may have been too coarse or homogeneous to differentiate between texts, or þulur as a genre may resist categorisation by social or regional variables. The presentation concluded with a reflection on the value of methodological experimentation even when results defy expectations, suggesting that such negative results are themselves significant and worth further exploration.

Ben Nagy (Institute for Literary Research of the Polish Academy of Sciences) presented "Syntax: The skeleton of style", arguing for the importance of syntactic features in stylometry. Historically, stylometry has focused on semantics, but scholars have recently demonstrated the merits of using syntax. Nagy used UDPipe to construct universal dependency trees from a corpus of sentences by five prose authors and six Latin poets, including Ovid. Using a tree classifier, he could correctly classify almost half of Ovid's sentences, while other authors' sentences were correctly classified only about a third of the time. To understand this syntactic distinctiveness, Nagy borrowed a tool widely used in model interpretability, using Shapley values to quantify and visualise the relative positive or negative contribution of each UD feature (drawn from relation type, head, POS and morphology) toward the binary classification of a sentence as Ovid or non-Ovid. AUX, the POS of auxiliaries such as the copula, emerged as the strongest contributor. Zooming in on Ovid, Nagy showed that it is possible to classify his verse into elegiac distichs and hexameters on the basis of syntax alone, corroborating the intuition that metre strongly constrains syntax and that caution is needed to avoid bias. Digging deeper, he further showed how the trees of Ovid's sentences fall into two loose groups along the x-axis when clustered in two dimensions. Dividing the plane into two halves and training a classifier to predict half, he found, for example, that one type prefers singular masculine nouns in the accusative case.

In her presentation "Keywords on Mainframes: What Can We Learn from Geir Kjetsaa's Studies in Computational Poetics?", *Antonina Martynenko* (Institute of Czech Literature, Czech Academy of Sciences) revisited the largely overlooked contributions of the Norwegian scholar Geir Kjetsaa (1937–2008) to the early development of computational poetics, situating his work within the broader institutional history of digital humanities and its canonical pioneers. Drawing on Kjetsaa's quantitative studies of Russian poetry from the 1970s and 1980s, she presented him as a figure who worked at the intersection of literary scholarship and early computing technologies to process and analyse large poetic corpora. The talk focused in particular on his 1973 study of Lermontov's poetic vocabulary and the creation of a "poetic norm" corpus

of 42,000 words from 19th-century Russian poetry, derived from a carefully selected canon of twenty poets. Martynenko showed that Kjetsaa was already engaging with issues that remain central to computational literary studies today, such as corpus construction, authorial representativeness, text sampling, and the problem of canonicity bias. At the same time, she critically discussed the limitations of his methodological framework, including his reliance on now outdated statistical measures. A key element of the paper was the replication of Kjetsaa's experiments using contemporary digital resources, including a reconstructed version of his original "norm" corpus. By comparing Kjetsaa's results for type–token ratios, keyword lists and vocabulary distinctiveness with those obtained through current techniques (including weighted log-odds ratio and Burrows' Zeta), the speaker demonstrated both the historical value and the limited reproducibility of his findings, especially with regard to keyword extraction. The presentation convincingly argued that Kjetsaa's work should neither be uncritically revived nor dismissed as obsolete, but rather reassessed as an important early attempt to formalise questions about poetic language with computational means. It concluded that historical computational studies can still be a valuable resource for contemporary research, provided they are approached critically and reinterpreted in light of modern data practices, theoretical frameworks, and current scientific standards, rather than simply replicated or imitated.

Mihhail Lotman (University of Tartu) presented his paper "Verses Hidden Behind Three Curtains", offering an interpretative and formally grounded analysis of Vladimir Nabokov's novel *The Gift* (1952), with a particular focus on embedded poetic structures. The paper centred on the novel's final sentence – "Eugene from his knees will rise" – which initially seems cryptic but is revealed to be part of a hidden stanza written in continuous verse. The stanza's structure corresponds to the Onegin stanza, evoking Pushkin's *Eugene Onegin* and inviting the reader into a multilayered intertextual dialogue. Lotman explored how Nabokov, writing under the pseudonym Sirin, attributed the novel to a fictional author, while its protagonist, Godunov-Cherdyntsev, is himself a poet. This complex authorial layering is further enriched in the fourth chapter, which includes Godunov-Cherdyntsev's fictional novel about Chernyshevsky. Though presented in prose, this embedded text contains poetic fragments – including versified citations from Marx – which did not originally exist in verse, raising compelling questions about internal authorship and the transformation of prose into poetry. These poetic fragments are presented as continuous text, blurring the boundary between verse and prose. Lotman emphasised the importance of metrical analysis, showing that Godunov-Cherdyntsev's poetic production is largely in iambic tetrameter, though the

rhythmic patterns vary between his early and mature works. This distinction, confirmed through statistical analysis, further illustrates the metafictional and metapoetic dimensions of Nabokov's novel.

Under the title "The Digital Lens: Irony Detection in Works of Juliusz Słowacki", *Anna Mędrzecka-Stefańska* (Institute for Literary Research of the Polish Academy of Sciences (IBL PAN)) examined the potential and limitations of digital methods for detecting irony and autothematic strategies in the poetry of Juliusz Słowacki, situating her study within broader discussions of Romantic irony as a mode that foregrounds the act of creation itself. Drawing on both theoretical accounts of Romantic autothematism and computational tools such as topic modelling and concordance analysis, she constructed two subcorpora of Słowacki's poems – texts classified as ironic and non-ironic – and analysed them with respect to the frequency and distribution of vocabulary linked to poetic creation, including terms denoting the poet, the act of writing, literary genres, and symbolic entities such as the Muse. While an initial comparison of word forms revealed only minimal differences between the two corpora, subsequent analyses of characteristic vocabulary and grammatical features, particularly verb forms associated with creative acts, uncovered more substantial divergences that nevertheless remained insufficient for unambiguous classification. Through examples drawn from *Journey to the Holy Land from Naples* and supported by wide-ranging textual evidence, Mędrzecka-Stefańska argued that irony in Słowacki's work emerges from the tension between creative intention and linguistic constraint, and from the polyphonic interplay of voices, perspectives, and citations that destabilise authorial authority. Overall, the presentation demonstrated that although digital tools alone cannot determine the presence of irony, they provide valuable insights into lexical tendencies, thematic clustering, and grammatical patterns that enrich traditional interpretative approaches and illuminate the intricate mechanisms through which Romantic irony operates in Słowacki's oeuvre.

Manex Agirrezabal (University of Copenhagen) discussed his ongoing research on the automatic scansion of Basque poetry in his presentation "Automatic Scansion of Basque Poetry: Many Questions and Few Answers (Yet)", approaching the topic from the perspective of computational linguistics and comparative metrics. Building on scansion models originally developed for English, he presented the design and manual annotation of a Basque poetry corpus consisting of approximately fifty poems and over 2,400 lines, with careful attention to authorial and chronological distribution. A substantial part of the talk was devoted to methodological issues, including text selection criteria, tokenisation, automatic morphological analysis, syllabification procedures, and the subsequent manual annotation of stress patterns.

Agirrezabal emphasised the specific challenges posed by Basque, a language in which the role of stress in poetic rhythm remains theoretically contested and where much modern poetry is strongly isosyllabic rather than accentual. His computational experiments demonstrated that, in contrast to English, more sophisticated machine-learning and neural approaches (including BiLSTM-CRF architectures) yield only marginal improvements over simple baseline methods for Basque, and often produce outputs that are difficult to interpret linguistically. These findings appeared to support the hypothesis that stress may not function as a primary organising principle of rhythm in much Basque poetry. Importantly, he also offered a self-critical reflection on the annotation process, openly discussing his doubts regarding the consistency of his own tagging and stressing the need for re-annotation and inter-annotator agreement in future work. Overall, the presentation provided a methodologically rigorous and self-reflexive account of the difficulties inherent in transferring computational scansion models to a less-resourced and typologically distinct language, and demonstrated how even partially inconclusive results can contribute to refining both theoretical assumptions and technical practices within computational poetics.

Hilofumi Yamamoto and *Bor Hodošček* (Institute of Science Tokyo, The University of Osaka) offered a detailed analysis of the mechanisms through which poetic thought is transformed into the highly constrained 31-syllable structure of classical waka in their talk "Transforming Poetic Thought into Waka: How to Pack the Skeleton into a 31-Syllable Closet". Using a parallel corpus of *Kokin Wakashū* poems and ten sets of modern Japanese translations, they investigated the compositional processes that enable poets to compress complex semantic and emotional content into the five-segment 5–7–5–7–7 form. By aligning original poems with their contemporary renderings through a hierarchical semantic annotation system (BG-codes), the speakers demonstrated that the characteristic condensation of waka relies on a repertoire of rhetorical and linguistic strategies, including the omission of subjects, predicates, and explicit markers of agency; the use of conventional symbolic substitutions (such as "bird" functioning as a metonym for "nightingale"); the inversion of canonical word order; and the deployment of nominalisation to create suspension and semantic density. Their residual analyses highlighted which elements translations must add – temporal adverbials, clarifying phrases, or abstract nouns – to make explicit what the original poems deliberately leave implicit, revealing a consistent preference in waka for concrete imagery, moment-specific seasonal references, and compressed Sino-Japanese compound structures over explanatory description. On this basis, Yamamoto and Hodošček proposed that waka composition adheres to

a "process grammar" grounded in immediate expression and iterative adjustment rather than standard Japanese syntactic norms. They argued that this grammar reflects both culturally specific aesthetic values and more universal poetic strategies for managing meaning under strict formal constraints. The presentation thus provided a richly documented account of poetic compression in waka and demonstrated the value of hands-on, explainable comparison techniques for uncovering the linguistic and cognitive operations that shape premodern Japanese poetics.

"Typology and Evolution: Revisiting the History of Romanian Poetry with Computational Tools" was the title of a joint presentation by *Andrei Terian*, *Mihai Dascalu*, *Aura Cristina Udrea*, and *Iuliana-Loredana Bîrsan* (Lucian Blaga University of Sibiu and the National University of Science and Technology Politehnica Bucharest). The presentation addressed the question of whether the traditional classifications of Romanian poetry – widely used by literary historians – could be confirmed through computational methods. The team compiled a diachronic corpus of 500 poems, representing ten poetic paradigms from 1830 to 2025, each paradigm being illustrated with texts by five canonical poets. These paradigms were drawn from influential Romanian literary histories and included categories such as Pre-Romantic, Romantic, Modernist, Traditionalist, Avant-Garde, Socialist Realist, Neomodernist, Postmodern, Metamodernist, and Post-2K poetry. To test the stylistic coherence and distinctiveness of these categories, the poems were analysed with the ReaderBench framework, a multilingual computational tool originally designed for prose texts. The team adapted it to verse by redefining structural units, enabling a detailed computation of over 200 textual complexity indices across multiple linguistic dimensions (surface, syntactic, semantic, and discursive). The results allowed the researchers to revisit longstanding historiographic questions, such as whether these categories hold up under quantitative scrutiny, how they cluster in relation to each other, and where significant paradigm shifts occurred in the evolution of Romanian poetic style. The study exemplified the methodological potential of computational tools in reinterpreting literary periodisation.

In their collaborative presentation "Do Zodiac Signs Influence Authorial Style? No.", *Petr Plecháč*, *Artjoms Šeļa* (both of the Institute of Czech Literature, Czech Academy of Sciences) and Neža Kočnik (University of Maribor) introduced a series of stylometric experiments designed to test a playful yet methodologically instructive hypothesis: whether an author's zodiac sign leaves any detectable trace in poetic style. Using the multilingual PoeTree corpus as their primary dataset, they applied a wide range of computational classification techniques to determine whether zodiac signs could be predicted

solely on the basis of stylistic features. Across the complete corpus, accuracy rates hovered around 8%, effectively identical to random chance and offering no evidence of astrological influence on poetic diction or rhythm. Separate analyses of individual language subcorpora yielded similarly null results for Czech poetry, confirming that stylistic variation is not patterned according to astrological categories. Only in the Russian subcorpus did the models achieve somewhat higher accuracy; however, as the speakers emphasised, this was not evidence of hidden cosmic determinism but rather a confounding effect of historical demography and canon formation. Because certain zodiac signs – such as Cancer, associated with late June and July – were overrepresented among canonical Russian poets of the late nineteenth century, the classifier effectively learned chronological and canon-based clustering rather than any stylistic properties linked to birth dates. This imbalance, humorously described as a "sweet summer child" effect, even revealed historical fluctuations in seasonal birth peaks, undermining the assumption of stable zodiac distribution. Overall, the results demonstrated not only the complete absence of correlation between zodiac signs and poetic style, but also the ease with which spurious patterns can arise from demographic or editorial biases, underscoring the importance of critically examining metadata and corpus construction when interpreting stylometric outcomes.

In "Romance Syllabics and Slavic Tonics in the 20th Century: Unexpected Influence", *Vera Polilova* (Lucian Blaga University of Sibiu) re-examined the origins of tonic and tonico-syllabic forms in early 20th-century Russian poetry, challenging the traditional emphasis on Russian folk verse and German models by arguing for a significantly underappreciated influence of Romance syllabic versification. Focusing on the period of metrical experimentation that culminated in the emergence of dolnik and related tonic forms, she demonstrated that Russian poets did not merely imitate the syllable counts of French, Italian, and Spanish poetry, but rather sought functional rhythmic equivalents within the non-isosyllabic structure of Russian verse. Drawing on a corpus of translations and imitations by poets such as Konstantin Balmont, Osip Mandelstam, Nikolai Gumilyov, and Valentin Parnakh, she showed how alternating intervals of different syllabic lengths within a line – perceived in Romance traditions as rhythmically salient – were reinterpreted in Russian tonico-syllabic terms through variable inter-ictic spacing. A central case study concerned Alexander Blok's *Verses on the Beautiful Lady* (1905), which she discussed against the background of existing interpretations by scholars such as Viktor Zhirmunsky and Roman Jakobson, and for which she proposed new Italian prototypes for two of the metres used in the collection. This reinterpretation questioned the view of Blok's dolnik as a purely internal evolution of Russian versification and

instead situated it within a broader European metrical dialogue. By carefully juxtaposing Romance syllabic originals with both isosyllabic and non-isosyllabic Russian adaptations, Polilova convincingly argued that what might appear as "pure tonic" innovation was in fact shaped by cross-linguistic rhythmic perception and translation practice. The presentation thus contributed to a more nuanced, transnational understanding of modern Russian versification, highlighting the extent to which early 20th-century poetic innovation emerged from complex interactions between different European metric systems rather than from isolated national traditions.

The presentation titled "Limericks in the Closet: The Unseen World of Estonian Limericks" by *Maria-Kristiina Lotman* and *Rebekka Lotman* (both of the University of Tartu) explored the largely unpublished and digitally circulated corpus of Estonian limericks. While the limerick is a well-established form in English poetry, its adaptation in Estonian literature has remained peripheral, often excluded from formal publishing and confined to private collections or social media platforms. The paper outlined the historical trajectory of Estonian limericks, from their first emergence via a poetry competition in the Swedish-Estonian newspaper *Vaba Eesti* in 1955, through their later development in Estonian periodicals during the 1970s and 1980s, and their incorporation into children's poetry only recently. Drawing on a database of printed limericks, the authors observed that the form predominantly adhered to the canonical five-line structure with an AABBA rhyme scheme and a pattern of three stresses in lines A and two in lines B. The study focused in particular on the unpublished limericks of the poet Ilmar Laaban, who introduced sound poetry to Swedish and Estonian literature and formulated a poetics based on the breaking of three taboos – universal, societal, and personal. Selected limericks from Laaban's personal archive revealed both formal inventiveness and taboo-challenging humour. The folklorist Madis Arukask also featured as a prominent example of contemporary, unpublished limerick writing, sharing his poems exclusively on Facebook. The paper concluded by discussing self-censorship as a key factor limiting the limerick's publication, both at the level of individual poets and within editorial processes.

In "The Echoes of the Noise or the Desperate Hope to Understand the Sound of Silence", *Andor Horváth* and *Levente Seláf* (from Eötvös Loránd University (ELTE)) reflected on the challenges and partial failures of a four-year research project dedicated to the analysis of 16th-century Hungarian epic poetry. Focusing on the methodological difficulties of dealing with a highly heterogeneous corpus, the authors tested a range of computational and philological approaches, including rhyme structure analysis, detection of syntactic parallelisms, stylometry based on morphological structures, rhythmic

modelling in search of a shared stanzaic origin, and morphosyntactic profiling. Despite analysing 175 epic poems – amounting to more than 99,000 lines – and producing extensive graphs, tables, and statistical visualisations, many of the resulting patterns ultimately proved either misleading or inconclusive. Morphological sequence analysis yielded dendrograms devoid of interpretative value, while attempts to explain metrical phenomena through prosodic evolution lacked statistical support. Repetition analysis initially suggested traditional or oral structures, but many of the matches could be traced back to idiomatic expressions, layout errors, or segmentation inconsistencies. Nevertheless, the project did lead to some unexpected yet philologically relevant discoveries, such as newly identified acrostics and oral traces in seemingly learned texts. Through a transparent account of both failed and fruitful methodologies, Horváth and Seláf offered valuable guidance for future computational and quantitative studies of historical verse traditions, highlighting the importance of corpus curation, theoretical restraint, and methodological flexibility.

Lidia Pivovarova (University of Helsinki), *Eetu Mäkelä* (University of Helsinki), *Kaarel Veskis* (Estonian Folklore Archives, Estonian Literary Museum, Institute of Estonian and General Linguistics, University of Tartu), *Kati Kallio* (Finnish Literature Society & University of Helsinki), *Mari Väina* and (Estonian Literary Museum) *Jakob Lindström* (University of Helsinki) presented "Evaluating Large Language Models for Translation and Annotation of Finnic Runosongs: Methodological Challenges". Runosong is a traditional trochaic-tetrameter song widespread in the Finnic languages and also used in literary works, for example, in the *Kalevala*. The corpus used by the authors consists of around 250,000 texts in seven different languages, recorded between 1564 and 1960. Part of the difficulty of doing NLP in these languages is the morphological richness: the corpus contains 1,084,389 unique word forms. Several of the languages, such as Karelian, Ingrian, Votic, Livonian and Veps, are small and extremely low-resource. Daunting for human annotators, this large corpus provides an ideal testing ground for assessing the abilities of large language models and how prompt engineering can improve them. The authors investigated four NLP tasks: on the word level, English translation, normalisation, lemmatisation and finding the etymological root, and on the level of an entire poem, dialect, orthography and genre. A test set of two hundred poems was manually annotated. DeepSeek and Claude both achieved over 90% on normalisation and lemmatisation but only 63% on etymology (all three measured as character-level Levenshtein distance), and 70% on translation (measured as cosine similarity between word embeddings). Interestingly, the smaller Llama 3 performed worse overall but slightly better on etymology. Problems with the models standardising words instead of normalising

orthography within the dialects led to making the prompt more explicit and including a detailed "morphological reconstruction protocol". This effort yielded improved scores across the board, highlighting the value and potential of prompt design optimisation.

Under the title *"(A)versions: Comparing Performances of the 'Same' Poem"*, Plínio A. Barbosa (University of Campinas, Brazil), Stefan Blohm (IDS Mannheim, Germany), and Chris Mustazza (University of Pennsylvania) explored the performative variability of poems by analysing multiple recordings of poets reading their own work over time. Drawing on the extensive PennSound archive – which houses over 65,000 poetry recordings – the team focused particularly on Robert Creeley's "I Know a Man", for which thirteen performances spanning four decades are available. Through a combination of phonetic, prosodic, and perceptual methods, the researchers examined how specific acoustic features of Creeley's readings evolved. Using the Prosody Descriptor Extractor script, they extracted 22 parameters related to pitch, voice quality, intensity, and tempo, enabling detailed comparison of changes both across the poet's lifetime and within individual poems. Results indicated patterns consistent with vocal ageing, such as slowing tempo and shifts in voice quality, as well as expressive variation tied to specific lines – most notably the line containing the phrase "goddamn big car", which showed marked prosodic deviation. Further, the shift between first- and third-person speech ("I said" to "he said") was accompanied by increased vocal effort and tempo, suggesting perceptible changes in narrative perspective. This work challenges fixed notions of a poem as a static text, foregrounding instead its status as a dynamic and temporally embedded performance event.

Anastasia Belousova (Universidad Nacional de Colombia) addressed the application of quantitative methods in translation studies, particularly in the context of Russian translations of poetry, in her presentation "Violets, Crucible, Measurement: Quantifying Features in Poetic Translation". While versological approaches traditionally focus on rhythm, rhyme, and syntactic structure, Belousova extended these tools to the evaluation of translated texts. Drawing on earlier work by scholars such as Gasparov, Polilova, and Galkin – who quantified aspects like lexical accuracy, rhythmic fidelity, and stylistic coherence – she examined how measurable parameters can be used to compare translations that have been differently received in literary criticism. The presentation explored translations of poetry from English, Spanish, and Italian into Russian by figures such as Konstantin Balmont, Samuil Marshak, Alexander Ilyushin, and Mikhail Gasparov. Belousova focused on translations that have been criticised or labelled as "poor", subjecting them to quantitative analysis in terms of lexical selection, rhythmic structure, and other stylistic

markers, and contrasting them with more favourably received versions. Her findings emphasised both the potential and the limitations of such methods: while they offer concrete tools for describing translation strategies, they often provoke resistance within the field of literary translation studies, especially when seen as implicitly evaluative. Nonetheless, the presentation highlighted that even this resistance can be productive, forcing a critical reflection on the role of subjectivity in translation reception and the epistemological stakes of quantification in literary studies.

The final day of the conference opened with *Albin Thörn Cleland's* (Lunds universitet) presentation, "Hidden Choral Stimuli: The Role of Accent in Aristophanes' Refrains". His paper examined the interplay of pitch accent, metre and musical melody across metrically identical refrains in pre-Hellenistic Ancient Greek lyric poetry, an aspect that has long remained relatively obscure and misunderstood. Building on Conser (2020), who first operationalised the link between the orthographic surface and the theory of music and prosody ("melodic compatibility") through open-source software, Thörn Cleland has developed a new framework that extends this functionality. Whereas Conser worked with the antistrophic two-refrain songs of tragedy, the new software is bespoke to the multi-refrain songs of comedy. Drawing on a corpus of all 79 multi-refrain songs in Aristophanes, Thörn Cleland performed three statistical tests. The first, a chi-square significance test, introduced two methodological refinements over Conser's approach: (1) using lyric tetrameter baselines rather than iambic recitative, and (2) generating one hundred randomised baselines instead of only a few. This test showed that only the two-refrain songs reached significance ($P = 0.0001$). The second analysis, a linear regression of the songs ordered chronologically, indicated no diachronic trend in melodic compatibility – contrary to Conser's findings for Aeschylus. Finally, Thörn Cleland demonstrated that the final syllable of each verse line exhibits exceptional melodic compatibility, even though no overall linear tendency is observed.

In their presentation "Machine Poetics: Music Genre Classification Using Prosodic, Stylistic, Syntactic, and Sentiment-based Features", *Stefan Baghiu* and *Erik-Robert Kovacs* (Lucian Blaga University of Sibiu) discussed their large-scale computational study of approximately 14,000 Romanian song lyrics, drawn from a wide range of publicly available online sources. Using natural language processing techniques, the researchers examined how song lyrics differ across genres based on prosodic, syntactic, stylistic, and sentiment-based features. To support this, they developed an improved lexical analysis tool tailored to Romanian poetic and lyrical language, encoding seventeen linguistic features and enhancing sentiment detection through the manual rebalancing of the SentiLex lexicon. These refinements yielded a 25% accuracy increase

over the original tool, as validated on a benchmark Romanian sentiment dataset. Their analysis revealed that genres traditionally seen as oppositional – such as rock and religious music – often exhibited striking formal similarities, particularly in repetition patterns and lexical structure. Hip-hop emerged as the most linguistically complex genre, while folklore and *manele* featured simpler syntax and stronger sentimental polarity, often oscillating between positive and negative affect. Baghiu and Kovacs emphasised that these results challenge entrenched aesthetic and cultural hierarchies, revealing unexpected structural affinities and pointing toward the need for more nuanced genre modelling. They concluded by outlining plans to expand their analysis to include 19th-century Romanian poetry, using the MDPR corpus developed at Lucian Blaga University, thereby bridging contemporary music studies with historical literary research in the digital humanities.

In his presentation titled "Aspects of Humour Salient to Large Language Models: Says Who?", *Pablo Ruiz Fabo* (Université de Strasbourg & CiTIUS – Universidade de Santiago de Compostela) extended the line of research he introduced at the previous year's conference, focusing on how large language models (LLMs) identify and explain humour in poetry. Building on earlier work that used OpenAI's non-specialised models to classify 100 Spanish sonnets (half humorous, half non-humorous), Ruiz Fabo replicated the experiment using newer models, such as GPT-4o-mini and both small and large versions of Mistral, besides Claude, Gemini and DeepSeek. These models were tested in zero-shot classification tasks and were prompted to generate short continuations and interpretive justifications for their decisions. Although performance metrics (especially F1 scores) showed notable improvement in comparison to earlier LLMs – Claude models reached a particularly high level of accuracy – key challenges persisted. Among these were the frequent misidentification of parody and subtle irony, as well as occasionally overconfident false positives. A novel aspect of this year's study was the use of supervised learning to determine whether it is possible to identify which LLM generated a given humour explanation. The presentation highlighted the latent stylistic and conceptual biases embedded in each LLM's understanding of humour. Ruiz Fabo concluded by emphasising that automatic humour detection remains a complex task – one that reveals more about the models themselves than about the poems they analyse.

The conference closed with the presentation "JIGS of Unburied Figures", in which *Anne-Sophie Bories* and *Nils Couturier* (both of the University of Basel) introduced a novel annotation framework – the Joke-like Incongruity Gathering System (JIGS) – developed within the project *Le Rire des Vers (Mining the Comic Verse)*. JIGS builds on Raskin and Attardo's General Theory

of Verbal Humour, particularly the concept of script opposition, yet shifts the focus away from overtly humorous texts toward literary corpora where incongruity is structurally present but not necessarily intended to elicit laughter. Applied to a corpus of French poetry and song lyrics (including works by Apollinaire and Renaud), the system allows annotators to identify segments marked by double-script incongruities – semantic tensions or reversals that function similarly to joke punchlines. The analysis includes a qualitative comparison of expert and non-expert annotator agreement, revealing substantial variation depending on the text and highlighting methodological challenges in identifying and labelling "joke-like" structures. Examples range from semantic inversions in Apollinaire's *Zone* to metaphorical density in Renaud's *Les Charognards*, with recurring oppositions like life/death, high/low status, and human/non-human. By extending humour theory to a broader poetic context, the authors explore the potential of JIGS as a tool for capturing semantic density and tension in verse. The talk reflects on annotation failures and unexpected readings as productive entry points for refining the model, ultimately arguing for a plural, interpretation-sensitive approach to operationalising literary incongruity.