Veras Audire et Reddere Voces: A Corpus of Prosodically-Correct Latin Poetic Audio from Large-Language-Model TTS

Michele Ciletti[1,*]

[1]University of Foggia, Department of Humanities, 71121 Foggia, Italy

Abstract

Latin verse moves to the pulse of vowel quantity and stress, yet students and researchers still struggle to hear that rhythm because high-quality recordings are rare and expensive. General-purpose text-to-speech models, rich in English and Romance data, flatten the long–short alternation that defines classical metres. This paper introduces a fully open, expertly validated corpus that demonstrates how far prompt engineering alone can push contemporary large language models toward metrically faithful Latin speech.

Drawing on Pedecerto's XML scansions, two emblematic passages were selected: the first 100 verses of Vergil's *Aeneid* (pure hexameter) and the opening elegiac *epistula* of Ovid's *Heroides* (elegiac couplets). Each line was syllabified, marked for ictus, elided where compulsory, and orthographically nudged into forms modern TTS engines pronounce reliably. The pre-processed verse, printed verbatim inside a concise system prompt, was rendered ten times by `gpt-4o-mini-tts`; human experts in Latin phonology then audited every take for segmental accuracy, stress placement, elision, and pacing. The accepted files were loudness-normalised and concatenated with uniform verse pauses, yielding roughly 24 minutes of continuous yet metrically autonomous recitation.

The release bundles (1) the original Pedecerto XML, (2) classical and TTS-ready transcriptions, and (3) per-line and stitched mp3 audio, all under CC-BY 4.0 and archived on Zenodo. Beyond serving as classroom audio or accessibility material, the aligned data provide a test-bed for prosody-aware speech synthesis, few-shot fine-tuning, and quantitative metrics research. An analysis of error patterns, such as cross-lingual accent drift, cluster mispronunciation, and length-stress trade-offs, offers concrete heuristics for steering future models without costly retraining. The workflow, implemented entirely with easily accessible APIs and lightweight scripts, could be readily transferable to Ancient Greek, Classical Arabic, or any verse tradition equipped with digital scansion. In short: the dead can be made to speak - rhythmically, reproducibly, and in the open.

Keywords

Latin, prosody, dataset, poetry, text-to-speech

1. Introduction

Latin verse obeys laws of rhythm that differ markedly from those governing most modern poetry. Whereas English metres depend on patterns of stress, classical Latin organises lines around the alternation of long and short syllables [1]. The term prosody itself stems from Greek *prosoidia*, which first referred to a tune sung to music, then to the pronunciation of individual syllables.

Convincing spoken performances of Latin poetry that students can consult remain remarkably scarce. Grammars and handbooks describe reconstructed pronunciations with care, still few recordings reproduce the quantitative rhythm that defines classical metres. Recent advances in neural text-to-speech have brought modern languages to broadcast quality, but Latin has been left on the margins: general-purpose models have little or no training data and therefore transfer English or Romance stress patterns almost wholesale.

The previous generation of text-to-speech (TTS) systems, exemplified by architectures like `Tacotron 2` [2], typically followed a two-stage process, converting text to an intermediate representation (a mel spectrogram) before a separate vocoder synthesized the audio. While capable of high-fidelity output, these models required extensive, language-specific training data, leaving low-resource languages like Latin underserved. However, the emergence of large language models (LLMs) with direct audio-generation capabilities offers a new paradigm. Unlike earlier architectures that required costly fine-tuning, these models can be steered through carefully crafted prompts. Models such as `gpt-4o-mini-tts` [3] employ end-to-end architectures that learn from vast, multilingual datasets and can be conditioned directly through in-context learning. This prompt engineering approach allows for fine-grained control over pronunciation, pacing, and emphasis without retraining the model, opening a promising avenue for generating prosodically-correct speech in low-resource languages. To demonstrate the viability of this approach, this paper introduces *Veras Audire et Reddere Voces*, a fully open and expertly val-

CLiC-it 2025: Eleventh Italian Conference on Computational Linguistics, September 24 — 26, 2025, Cagliari, Italy

*Corresponding author.

michele.ciletti@gmail.com (M. Ciletti)

https://www.researchgate.net/profile/Michele-Ciletti (M. Ciletti)

0009-0004-3829-8866 (M. Ciletti)

idated corpus of prosodically-accurate Latin poetic audio. The corpus contains 216 lines of verse - the first 100 hexameters of Vergil's *Aeneid* and the opening 116 lines (58 elegiac couplets) of Ovid's *Heroides* - providing a total of nearly 24 minutes of metrically autonomous recitation. The primary contributions of this work are threefold: the release of a high-quality, aligned dataset of Latin poetic audio, TTS-ready transcriptions, and metrical annotations under an open license; the presentation of a reproducible workflow that shows how prompt engineering alone can steer a general-purpose LLM toward metrically faithful speech, offering a low-cost alternative to model retraining; and the production of a valuable resource for pedagogy and a test-bed for future research in controllable speech synthesis for historical languages.

2. Theoretical Background

2.1. Latin Prosody

Classical verse draws its rhythm from vowel quantity and from the consonantal context that can lengthen a syllable [4]. Six-foot dactylic hexameter and the coupled hexameter–pentameter of the elegiac distich are the metres most familiar to students. Rules such as *muta cum liquida* allow optional resolution; pervasive elision removes a vowel at word boundaries; and the location of the *caesura* shapes phrasing. Because no recordings survive from antiquity, quantity must be inferred from orthography, comparative Romance evidence, metrical practice, and statements by ancient grammarians. Absolute certainty is impossible, which explains why modern classrooms often substitute a stress-based reading, even though Latin stress itself follows a moraic algorithm. Any speech-generation system has therefore to decide which principle it will privilege: quantity, stress, or a compromise.

2.2. Digital Latin Resources

Over roughly thirty years Latin has acquired a considerable amount of Natural Language Processing resources [5] [6]. Tokenisers, lemmatisers, and treebanks are distributed through CLTK [7], Stanza [8], and Universal Dependencies [9]. Prosodic annotation is rarer. PedeCerto [10] marks quantity, feet, and *caesurae* for more than 240,000 dactylic lines; its XML export underlies the corpus described here. Other scanners cover particular metres: for instance CLTK modules for hexameter and hendecasyllable, Anceps for Senecan trimeter [11], or Loquax for syllabification and IPA transliteration [12].

2.3. Prompt-based Prosody in Large Language Models

Large language models that decode speech directly from text have begun to internalise prosodic patterns. Architectures such as `VALL-E` and `ZM-Text-TTS` train on vast multilingual collections; their outputs preserve speaker identity and sentence melody, yet metre remains hard to control [13]. One pragmatic strategy is to preprocess the poem itself: mark ictic syllables with capital letters and diacritics, resolve compulsory elisions, and substitute unfamiliar graphemes with spellings that the model already pronounces reliably (chiefly English, with Italian conventions for a selection of elements). During synthesis these visible cues bias the duration and stress predictors without any retraining of the model. The approach follows PRESENT's [13] principle of steering prosody through input representation rather than through explicit feature vectors.

2.4. Pedagogical and Inclusive Perspectives

High-quality recordings produced by trained classicists are time-consuming and costly. Automatic generation, once trustworthy, would make spoken Latin more accessible in schools, in digital humanities research, and for visually impaired learners. Surveys in the field call for FAIR corpora that combine text, audio, and metadata [14]. By publishing aligned verse–audio pairs, the present work answers that demand in part. Stress-centred recitation also lowers the entry barrier for students whose native languages do not contrast vowel length, while still preserving a perceptible rhythmic pulse consistent with traditional metrics.

3. Methodology

3.1. Source Texts and Metrical Gold Standard

The audio that accompanies the dataset was derived from two chosen passages: the first hundred hexameters of Vergil's *Aeneid* and the opening elegiac epistle of Ovid's *Heroides*. These segments supply, on the one hand, a pure run of dactylic hexameter and, on the other, the alternation of hexameter and pentameter typical of the elegiac couplet. Machine–readable scansion was taken from the XML export of the Pedecerto project [10]. Each `<line>` element preserves the metrical category, the canonical foot pattern and, for every word, a `sy` attribute that enumerates syllables while marking the ictus with an upper-case character. The import script retained verse boundaries, foot sequence, ictic flags, elision hints and

word–boundary information; all other metadata were discarded. A fragment of the XML illustrates the structure:

```
<line name="1" meter="H" pattern="DDSS">
  <word sy="1A1b" wb="CF">Arma</word>
  ...
  <word sy="2c3A" wb="CM">cano,</word>
  <word sy="3T4A" wb="CM">Troiae</word>
  ...
</line>
```

3.2. Pre-processing and Orthographic Adaptation

Each verse underwent an iterative pipeline before it was ever passed to the speech engine. Syllabification relied on the rule-based module distributed with the Classical Language Toolkit [7]; diphthongs and enclitics are already covered in that implementation. The vowel of every ictic syllable received a grave accent and the complete syllable was converted to capitals. Obligatory elisions were realised as graphic mergers (*quoque et* therefore became *quoquet*) according to the Pedecerto `wb` flag. A comma was inserted where the metre demands a caesura unless the manuscript already offered punctuation at that position. Early trials separated syllables with hyphens, but the additional markers produced no audible advantage and the idea was dropped.

A second pass substituted graphemes that tend to mislead English-trained acoustic models. Before front vowels ⟨c⟩ was rewritten as ⟨k⟩, ⟨qu⟩ became ⟨kw⟩, the diphthongs ⟨ae⟩ and ⟨oe⟩ were rendered ⟨ai⟩ and ⟨oi⟩, and palatal ⟨g⟩ was expanded to ⟨gh⟩. The resulting string approximates a classical pronunciation yet stays within the alphabetic habits of contemporary TTS systems.

To keep prosodic control local, each line was synthesised in isolation; the rhythm inside a verse must be coherent, whereas a small pause between verses is both acceptable and expected in performance.

3.3. Speech Generation and Iterative Refinement

Two technological families were explored. Conventional sequence-to-sequence TTS engines, such as `Tacotron 2` [2], `Kokoro` [15], OpenAI's `tts-1-hd` [16], offer little room for instruction: stress was frequently misplaced and vowel length erratic, particularly when the Latin token resembled a common English form. Multimodal large language models with an integrated audio decoder fared better because the system prompt can be used to impose a prosodic policy. Several models in the GPT-4o and Gemini lines were evaluated; `gpt-4o-mini-tts` [3] delivered the most consistent timing and segmental clarity.

Prompt engineering began with an extensive style sheet, eventually distilled to three imperatives: speak slowly, articulate every syllable, and obey the marked stresses. Re-printing the fully processed verse inside the system prompt, exactly as it should be spoken, noticeably improved alignment between text and realisation. Because stochastic sampling introduces variation, ten readings were requested for every line. The final system prompt was:

> This is a Latin poetical verse. Pronounce it rhythmically, slowly and with emphasis, articulating each syllable and correctly stressing them. Pronounce it like this: [pre-processed verse]

The addition of the word "slowly", explicitly telling the model to recite the verses at a relaxed pace, proved to be particularly useful in ensuring that each syllable was correctly articulated.

3.4. Human Validation and Error Annotation

Specialists in Latin phonology audited every recording. Errors were marked on spans and classified as segmental substitution or ictus misplacement. Feedback after each experimental round guided small adjustments to the pre-processing routine and to the wording of the prompt. Acceptance was granted when a line contained no error of stress or elision and no more than minor segmental deviations; under this criterion at least one satisfactory rendition was eventually found for each of the autonomous lines.

3.5. Mastering and Packaging

For every verse the reviewers selected the highest-scoring file. Selected waveforms were loudness-normalised and concatenated with an 800 ms silence, yielding two continuous recitations that preserve per-line rhythmic autonomy. Alongside the audio the repository contains:

- the original Pedecerto XML fragments,
- the full text of the chosen passages,
- the pre-processed lines which have been given as input to the TTS model.

All artefacts are released under an open licence and have been deposited on Zenodo together with a DOI, ensuring long-term accessibility and citability [17].

For a more thorough discussion of the methodological choices, from model selection to human evaluation, the reader is referred to a previous publication [18].

Table 1
Overview of the corpus

Sub-corpush	Metre	Lines	Hexameters	Pentameters	Total duration (hh:mm:ss)
Aeneid 1.1-100	Dactylic hexameter	100	100	0	00:11:26
Heroides 1.1-116	Elegiac couplet (hexameter + pentameter)	116	58	58	00:12:26
Total	-	216	158	58	00:23:52

4. Results: Description of the Released Corpus

The outcome of the workflow is an aligned collection of Latin poetic audio accompanied by the textual and metrical information required for downstream work in speech technology, pedagogy and quantitative metrics. The repository is organised around three sections: for each poem, a text file contains its original lines, another one has the pre-processed text that was fed to the TTS model, an XML file contains the Pedecerto metrical annotations and a set of mp3 files represent the audio output, stored both individually and as groups.

Table 1 gives an overview of the material.

4.1. Audio Layer

For each verse ten independent readings were decoded. After expert screening one rendition was retained as the canonical file. Recordings are stored as mp3 files. Silences at verse boundaries have been standardised to 800 ms; no fades or noise reductions were applied, so that the signal keeps its original spectral profile.

4.2. Text and Prosodic Annotation

The reference transcription follows the orthography employed during synthesis (grave accents on ictic vowels, upper-case ictic syllables, adapted spellings for *c, qu, g, ae, oe*) so that users can reproduce or extend the experiments without reverse engineering. A parallel file restores classical spelling for readers who prefer a diplomatic text. Pedecerto syllable scansions, foot divisions and caesura marks are displayed in a separate XML file.

4.3. Availability and Licensing

All components are released under CC-BY 4.0. The Zenodo record bundles the audio, textual content, and annotations [17].

5. Discussion

The present corpus was assembled in order to facilitate prosodically faithful speech synthesis, yet the labour invested in its creation has generated several observations that matter beyond the immediate goal of reciting Vergil and Ovid. Three strands of evidence stand out: the behaviour of the language model during synthesis, the practical tricks that secured acceptable output, and the prospective uses of the aligned data in research and teaching.

5.1. Accents, Cross-lingual Interference, and what the Model really “Knows”

When the decoder was left to its own devices it tended to interpret individual words through the accent template of whichever modern language offered the closest orthographic match. As a consequence, passages dominated by vocabulary shared with present-day Romance received an intonation reminiscent of Italian, while lines rich in loanwords familiar to English appeared with a markedly anglophone timbre. Spanish patterns surfaced less often, but, for example, whenever words ended in *-rant* the cadence was audibly Iberian. These drifts rarely broke the quantitative rhythm prescribed by hexameter or pentameter; they did, however, blur vowel quality, especially in the mid-front and mid-back zones. The phenomenon confirms that the model encodes a multilingual phonology for conversational prose, and stresses again how little purely Latin data the underlying training set must contain.

5.2. Problematic Phonotactics, Orthographic Workarounds and *Caesurae*

Several consonant clusters led to systematic errors. Final *-nx* in *coniunx* or initial *tl-* in *Tlepolemus* were clipped or resolved into epenthetic vowels, presumably because the sequences are rare in the speech material seen during pretraining. In other cases the word was recast according to a high-frequency modern homograph: *-um* often came out as *əm*, betraying an English proper-name template. Two heuristics mitigated these slips. First, lengthening the grapheme that carries the metrical ictus often persuaded the model to anchor stress correctly; *cano* became *caano* in the prompt, which silenced the temptation to

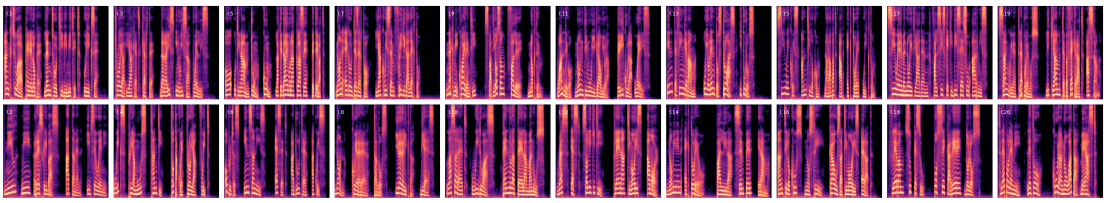

Figure 1: Mel Spectrogram of the first 20 lines of the opening *epistula* of the *Heroides*. Each tile corresponds to one metrical line; hexameters are at the top, while pentameters are at the bottom. Note how the pentameter's obligatory *caesura* after the third *arsis* is visible as a systematic pause mid-tile.

favour the English reading. Second, replacing rare digraphs with phonetically transparent ones, already documented previously, reduced segmental substitutions by a third. The interventions are admittedly ad hoc, yet they illustrate how a handful of hand-crafted rules can serve where large retraining runs are impossible. While improvements can certainly be made in regards to the overall flow of the generated verses, it remains one of the most effective features: the addition of commas to mark *caesurae*, in particular, proved to be useful in ensuring that the synthetic voices followed a precise pattern. Figure 1 shows the Mel spectrograms of the first twenty lines of the opening *epistula* of the *Heroides*, generated to visualize the verses' rhythm. The spectrograms clearly show that black bars (representing pauses) are always present in the middle of the bottom tiles (pentameters), while they are more spaced out in the top tiles (hexameters). This is due to the obligatory *caesura* that occurs after the third *arsis* of each pentameter, precisely in the middle of the verse, while hexameters present more varied structures.

5.3. From Corpus to Model Improvement

Because each verse is aligned with a verified audio, the collection can function as a fine-tuning set for both autoregressive and non-autoregressive TTS systems. A model trained on metrically correct examples should internalize the prosodic rules more reliably than a general-purpose system forced to extrapolate from modern language data. Even large language models themselves could benefit from exposure to annotated Latin verse during continued pretraining, potentially reducing the need for elaborate preprocessing in subsequent applications. Future work could examine whether freezing the acoustic front-end while training only the variance adapters suffices to introduce length contrast in addition to stress.

5.4. Classroom Impact and Reproducibility

In a teaching environment the recordings serve two complementary roles. Students can listen to a metrically regular rendition before attempting their own, and instructors can use the annotated text as input to alternative voices or slower tempos. Since every script that produced the audio leverages easily-accessible APIs, replication is straightforward. Such transparency matters especially for assessment settings where students must know exactly which variant counts as the reference.

5.5. Open Data and Transfer to Other Languages

Latin is only one among many historical or minoritised languages whose sound patterns are absent from mainstream speech technology. The workflow described here, licensed permissively and documented line by line, could be cloned for Ancient Greek, Old Occitan, Classical Arabic, or any verse tradition that already enjoys digital scansion. Riemenschneider and Frank [5] argue that large language models can be useful tools for Classical Philology; releasing small but expertly annotated sets therefore aims to accelerate progress. Open repositories also lower the entry cost for community contributors who may wish to supply alternative voices, extended passages, or corrected quantities.

5.6. Towards Length-sensitive Synthesis

Stress was easier to enforce than absolute vowel length. The present system approximates quantity indirectly through slower pacing on ictic syllables, yet it cannot keep a fixed ratio between heavy and light vowels. Lam et al. [13] report that explicit duration tokens unlock such control in English; integrating a similar mechanism with the current prompt-based strategy is an obvious next step. Ultimately, a synthesis pipeline that differentiates both stress and quantity would let classicists test competing reconstructions of Latin phonology "in silico", converting theoretical statements into audible hypotheses.

5.7. Outlook

Refining the preprocessing scripts, automating error spotting, and expanding the text base are immediate priorities. Nevertheless, even in its present form the corpus already supports experiments in few-shot prosody transfer, quantitative metrics, and accessible pedagogy. The value of such resources lies also in the demonstration that high-quality data can be gathered with modest equipment, provided that domain knowledge and iterative verification guide the process. Open, reproducible corpora therefore remain the necessary foundation on which future work for classical languages will build.

6. Limitations

The corpus was assembled with the intention of demonstrating what present-day language models can already achieve when prompt engineering is combined with careful human verification. Precisely because the focus lay on a proof of concept, several boundaries were accepted that restrict the scope of the resource. The most visible limitation concerns size. Only two passages, albeit canonical ones, entered the pipeline; together they furnish a little under twenty-four minutes of speech. For certain experiments in prosody transfer that duration suffices, yet quantitative studies of acoustic variance or full fine-tuning of an end-to-end synthesiser usually require at least an order of magnitude more material.

Closely related is the question of stylistic breadth. The *Aeneid* and the *Heroides* differ in metre, tone and lexicon, but both belong to the same literary period and represent the same formal register. Comedy, forensic oratory or Late Latin hymns remain untested. Consequently, the substitution rules that helped the model through epic and elegiac vocabulary might fail when confronted with colloquial forms, post-Classical spellings or heavy Greek loanwords.

Another constraint derives from the decision to rely on a single synthetic voice. Because speaker identity never changes, the corpus cannot inform studies that investigate how metre interacts with timbre or gendered pitch ranges. Similarly, only one variant of reconstructed pronunciation is encoded. Alternative schools that prefer the ecclesiastical pronunciation will find no examples that match their conventions. Validation, indispensable for quality control, introduces its own bias. Judgements about short hesitations or barely perceptible vowel colouring can differ across traditions; a panel of experts drawn from a wider set of institutions might have retained or rejected a slightly different subset of takes.

Technical choices add further caveats. Recordings were mastered to mp3 for ease of distribution, which entails lossy compression. The prompts are public, but the underlying model weights remain proprietary; should the provider change access policies, identical reproduction could become impossible. Finally, quantity was approximated through slower pacing on metrically strong syllables. The approach yields a rhythm that experienced listeners recognise, yet it falls short of enforcing a fixed heavy-to-light duration ratio, the gold standard in phonetic work on quantitative metres [4].

7. Conclusion

The study has introduced an openly licensed, line-aligned corpus that brings classical Latin verse within reach of modern text-to-speech technology. By combining Pedecerto's machine-readable scansion with a small set of orthographic substitutions and a concise prosodic prompt, the workflow coerced a general-purpose large language model into producing intelligible, metrically coherent recitations. Systematic human screening guaranteed that the released audio reflects the intended rhythm at a level suitable for both pedagogy and computational research.

The resulting dataset offers three immediate avenues of use. Teachers can deploy the files as accessible classroom material, learners may rehearse passages while receiving instant acoustic feedback, and speech engineers now possess a clean test bed for experiments in prosody conditioning. Beyond these practical gains, the project demonstrates that domain knowledge, when encoded explicitly in the input, still matters even in an era of ever larger pretrained models. Prompt design, although sometimes dismissed as a stop-gap measure, revealed itself here as a cost-effective alternative to full retraining.

Future work will have to broaden the metrical and generic range, increase speaker diversity and explore direct duration control. A longer term ambition is to fold the current resource into a multilingual library of verse corpora, so that comparative metrics across the Indo-European tradition become feasible. The dataset and annotations supplied with this release aim to render such extensions straightforward.

Acknowledgments

The author thanks the entire Pedecerto team for annotating and sharing their XML scansions. Sincere gratitude is extended to the CLTK community, whose open-source tools simplified syllabification and phonological checks. Colleagues at the University of Foggia donated hours to the auditory review of candidate recordings; their expertise shaped both the preprocessing rules and the acceptance thresholds. Any remaining inaccuracies are the sole responsibility of the author.

References

[1] B. W. Fortson IV, Latin prosody and metrics, A companion to the Latin language (2011) 92–104.

[2] J. Shen, R. Pang, R. J. Weiss, M. Schuster, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, R. Skerry-Ryan, R. A. Saurous, Y. Agiomyrgiannakis, Y. Wu, Natural tts synthesis by conditioning wavenet on mel spectrogram predictions, 2018. URL: https://arxiv.org/abs/1712.05884. arXiv:1712.05884.

[3] A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford, et al., Gpt-4o system card, arXiv preprint arXiv:2410.21276 (2024).

[4] W. S. Allen, Vox Latina: a guide to the pronunciation of classical Latin, Cambridge University Press, 1989.

[5] F. Riemenschneider, A. Frank, Exploring large language models for classical philology, arXiv preprint arXiv:2305.13698 (2023).

[6] B. McGillivray, Methods in Latin computational linguistics, volume 1, Brill, 2013.

[7] K. P. Johnson, P. J. Burns, J. Stewart, T. Cook, C. Besnier, W. J. Mattingly, The classical language toolkit: An nlp framework for pre-modern languages, in: Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing: System demonstrations, 2021, pp. 20–29.

[8] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, Stanza: A python natural language processing toolkit for many human languages, arXiv preprint arXiv:2003.07082 (2020).

[9] M.-C. De Marneffe, C. D. Manning, J. Nivre, D. Zeman, Universal dependencies, Computational linguistics 47 (2021) 255–308.

[10] E. Colombi, L. Mondin, L. Tessarolo, A. Bacianini, D. Bovet, A. Prontera, Pedecerto, Pedecerto. Metrica Latina Digitale (2011).

[11] A. Fedchin, P. J. Burns, P. Chaudhuri, J. P. Dexter, Senecan trimeter and humanist tragedy, American Journal of Philology 143 (2022) 475–503.

[12] M. Court, Loquax: Nlp framework for phonology, https://github.com/mattlianje/loquax, 2025. GitHub repository.

[13] P. Lam, H. Zhang, N. F. Chen, B. Sisman, D. Herremans, Present: Zero-shot text-to-prosody control, IEEE Signal Processing Letters (2025).

[14] M. De Sisto, L. Hernández-Lorenzo, J. De la Rosa, S. Ros, E. González-Blanco, Understanding poetry using natural language processing tools: a survey, Digital Scholarship in the Humanities 39 (2024) 500–521.

[15] Hexgrad, Kokoro-82m (revision d8b4fc7), 2025. URL: https://huggingface.co/hexgrad/Kokoro-82M. doi:10.57967/hf/4329.

[16] OpenAI, Openai tts-1-hd model documentation, 2025. URL: https://platform.openai.com/docs/models/tts-1-hd, accessed: 2025-06-29.

[17] M. Ciletti, Veras audire et reddere voces: A corpus of prosodically-correct latin poetic audio from large-language-model tts, 2025. URL: https://doi.org/10.5281/zenodo.15677356. doi:10.5281/zenodo.15677356.

[18] M. Ciletti, Prompting the muse: Generating prosodically-correct Latin speech with large language models, in: J. Zhao, M. Wang, Z. Liu (Eds.), Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop), Association for Computational Linguistics, Vienna, Austria, 2025, pp. 740–745. URL: https://aclanthology.org/2025.acl-srw.48/. doi:10.18653/v1/2025.acl-srw.48.

Declaration on Generative AI

During the preparation of this work, the author(s) did not use any generative AI tools or services.