

# Computer Science Project Contract

Leiden Institute of Advanced Computer Science  
Faculty of Science, Leiden University

## Agreement on Research Project/Master Thesis between:

### Student

**Name:** Luuk Nolden  
**Student number:** s1370898  
**E-mail address:** l.s.j.nolden@umail.leidenuniv.nl

### Supervisor(s)

**Name:** Suzan Verberne  
**Role:** First supervisor  
**E-mail address:** s.verberne@liacs.leidenuniv.nl

**Name:** Matthew Payne  
**Role:** Second supervisor  
**E-mail address:** m.f.payne@hum.leidenuniv.nl

---

## Research project

### Project summary:

The sound of a text is an essential part of its content. Similar to what we experience when we listen to different musical styles, our perception of a text depends on the rhythmical patterns in which this text is presented. Latin poetry contains lots of different rhythmical patterns, also known as metres, which can be seen as rules of sound and rhythm a line has to adhere to. However, these rules can be bent, subverted or be played with to create an atmosphere for the comedy, tragedy or poem.

Sadly, we cannot with certainty reconstruct all patterns: we do not always know whether a specific syllable had a bigger emphasis to convey magnitude, or was pronounced short to get a sense of urgency.

This project will investigate whether written Latin poetry inherently contains the information whether a syllable should be pronounced long, short or should be elided. As we know all the rhythm rules of the hexameter, epic texts like the Aeneid will be used as a ground truth to train our models on. Once a satisfactory model accuracy is reached, other rhythms like iambic metre and clausulae will be investigated to see if training a model on a hexameter will allow the scansion of more difficult metrical patterns.

Machine learning is an excellent way to try and achieve this, as it is more important to understand the principles by which the sounds of words create the rhythms than any specific (rule-based) patterns that may only work for one author or work. Just like a student needs to learn to apply these principles and to get a feeling for the relationship between these principles and how they work in practice in texts, so too for automatic scansion to succeed, the machine needs to be able to translate between these fundamental principles and specific metrical patterns as they are found in different authors and genres. This is why machine-learning approaches will be used, rather than trying to create a model that follows rules which only work for one kind of metrical pattern.

# Computer Science Project Contract

Leiden Institute of Advanced Computer Science  
Faculty of Science, Leiden University

This research is interesting for the study of Classics for helping students scan poetry and for helping researchers scan unknown metres. For the field of Computer Science it would be interesting to see whether rhythm is inherently encoded in written text and whether it could be retrieved using character, syllable and word embeddings.

**Working title:**

**Finding Sound and Rhythm in Written Latin Poetry**

**Date of project start:**

2021-09-17

**Deadline for handing in the report:**

2022-04-17

**Initial project plan / global planning:**

---

The global planning is as follows:

Sep 2021	Creating Naive models like Sequence Labeling and Hidden Markov Chains
Oct 2021	Designing Deep Neural Network: Bi-Word LSTM
Nov 2021	Creating word, syllable and character embeddings for Bi-Word LSTM
Dec 2021-Jan 2022	Testing with different texts, parameters and so on
Feb 2022	Finalize experiments and analyses, draft thesis
Mar-Apr 2022	Write thesis
Other agreements:	Meetings with first supervisor every 2 weeks, with 2 <sup>nd</sup> supervisor every week

---

# Computer Science Project Contract

Leiden Institute of Advanced Computer Science  
Faculty of Science, Leiden University

## Signatures

**Date:** 2021-10-06

**Location:** Leiden

**Student**

Luuk Nolden

**Supervisor**

Suzan Verberne

