

Objective Detection of Plautus' Rules by Computer Support

Deufert, Marcus

mdeufert@eaqua.net
Department of Classics, University of Leipzig,
Germany

Blumenstein, Judith

jblumenstein@eaqua.net
Department of Classics, University of Leipzig,
Germany

Trebesius, Andreas

atrebesius@eaqua.net
Department of Classics, University of Leipzig,
Germany

Beyer, Stefan

sbeyer@eaqua.net
Natural Language Processing Group, Institute
of Mathematics and Computer Science,
University of Leipzig, Germany

Büchler, Marco

mbuechler@eaqua.net
Natural Language Processing Group, Institute
of Mathematics and Computer Science,
University of Leipzig, Germany

The metre of the Roman comic poet Plautus (flourished ca. 200 B.C.) still leaves one mystified. Although the scientific work of the 19th and early 20th century has established a number of important rules and licences, the exact range of these laws and licences remains a matter of debate. Taking into account these many open questions it is not surprising that metrical studies (as well as the important editions) of Plautus still display a huge amount of discrepancy in their handling of Plautine metre. The specific problem consists of the large number of transmitted verses in the Plautine corpus and the great complexity and diversity of competing explanations of remarkable metrical phenomena.

Therefore, until now the results of scholarship often fail to convince, since they are based either on a limited textual basis or deal with a specific metrical phenomenon from the perspective of a

single law or licence without taking into account competing explanations.

This paper will cover a wide range of previous research from both Classics (Lotman 2000) and literature (Garzonio, 2006) to several techniques in the field of Computer Science (Heyer et al., 2008 and Volk, 2007). Metric analyses can be already be computed on German poems with only a small set of rules (Bobenhausen, 2009). Results imply, however, that foreign words are especially difficult to handle. In contrast to this, ancient texts pose a different problem as lots of variations of an original often exist. For this reason metric analysis can be divided into three different tasks:

- *Task 1: Dealing with different variations and variations of variations* (Andreev, 2009 and Rehbein, 2009). Within this paper, a primary version of a verse is defined by researchers from Classics. Differences of variants in relation to the primary version are highlighted as described in (Büchler et al., 2009¹ and Rehbein, 2009). The variance caused by transmissions of several authors is also important to consider, however, when working with fuzziness. Consequently a set of possible metric analysis annotations are suggested rather than just one result.
- *Task 2: Applying a metric rule-set* to a text corpus (Bobenhausen, 2009 and Fusi 2008). Within this research - similar to part of speech tagging (Heyer et al., 2003) – a set of rules is applied to text. However, only the most probable metric candidate is selected. In contrast to that research (Bobenhausen, 2009 and Fusi, 2008), the approach in this paper scores several possible metric analyses.
- *Task 3: Training of a metric ruleset* based on manually annotated data from researchers. Typically, a fixed set of rules is taken as presumed, however, new rules need to be added manually. This paper thus also focuses on the computation of new rules. The importance of this step is motivated by the Theory of Selective Perception. Based on this, new and uncommon rules are determined by a computer model that is both objective and independent rather than selective like a human being.

In the field of natural language processing the task of tagging text is quite similar to

part of speech tagging (POS). Typically, for such a tagger a Hidden Markov Model (Heyer et al., 2008) is trained and is traversed by dedicated algorithms like the Viterbi algorithm (Heyer et al., 2008). However, the already mentioned fuzziness of text variants makes both the training and traversing steps difficult. Furthermore, in the training step it is necessary to observe data on a larger window than the typical memory of 2 or 3. This would increase the complexity drastically during the trainings phase. Within the applying phase the Viterbi algorithm is typically used (Heyer et al., 2003). This algorithm reduces all paths locally except the most probable one. In metric analysis however this assumption is quite critical since due to syllable fusion an senarius is not required to have 12 but can also consist of 17 or 18 syllables.

Motivated by the aforementioned problems of existing approaches this paper describes a three step approach. In a first step possible syllables are computed. This is simply done by using training data. In contrast to German poems (Bobenhausen, 2009) the approach is aware of possible fusions of syllables. In the second step all possible combinations are computed instantly removing candidates that do not fulfil the metric requirements. The training itself is done by distance-based co-occurrences (Büchler, 2008) on metric tags. In the last step metric candidates are scored based on both the training data as well as the variance of the alternatively transmitted variances. All relevant candidates are selected by researchers of Classics. The remaining metric analysis is represented in a dedicated visualisation highlighting the differences of several variants to the primary version (Büchler, 2009 and Rehbein, 2009).

As an outcome of this paper several results will be shown. Besides the difference visualisation both results and experiences in training and application of a metric model will be provided.

Both the expected results and the developed software, which can be easily adapted to other ancient poets, will give an original input to the research community and motivate and enable further investigations in the same spirit.

References

- Andreev, V. S.** (2009). 'Patterns in Style Evolution of Poets'. *Digital Humanities 2009*. Pp. 52-53.
- Bobenhausen, K.** (2009). 'Automatisches Metrisches Markup'. *Digital Humanities 2009*. Pp. 69-72).
- Büchler, M.** (2008). *Elemente einer Forensischen Linguistik* Working report.
- Büchler, M., Geßner, A.** (2009). 'Unsupervised Detection and Visualisation of Textual Reuse on Ancient Greek Texts'. *2009 Chicago Colloquium on Digital Humanities and Computer Science*. Chicago, Nov. 2009.
- Fortson, B. W.** (2008). 'IV, Language and Rhythm in Plautus'. *Synchronic and Diachronic Studies*. Berlin / New York.
- Fusi, D.** (2009). *An Expert System for the Classical Languages: Metrical Analysis Components*. <http://www.fusisoft.it/Doc/ActaVenezia.pdf><http://www.fusisoft.it/Chiron/Metrics/Default.aspx> (accessed Nov. 10th 2009).
- Garzonio, S.** (2006). 'Italian and Russian Verse: Two Cultures and Two Mentalities'. *Studi Slavistici*. **III**: 187-198.
- Heyer, G., Quasthoff, U., Wittig, T.** (2008). *Text Mining: Wissensrohstoff Text – Konzepte, Algorithmen, Ergebnisse*. 2nd edition. W3L-Verlag.
- Lotman, M.-K.** (2009). 'Word-ends and Metrical Boundaries in Ancient Iambic Trimeter of Comedy'. *Studia Humaniora Tartuensa*. **1**: 1-16. <http://www.ut.ee/klassik/sht/2000/lotman1.pdf> (accessed Nov., 10th 2009).
- Questa, C.** (2007). *La metrica di Plauto e di Terenzio*. Urbino.
- Rehbein, M.** (2009). 'Multi-Level Variation'. *Digital Humanities Conference Abstracts*. 2009 2009, pp. 11-12.
- Volk, A.** (2007). 'Rhythmic Similarity based on Inner Metric Analysis'. *Utrecht Summer School Multimedia Retrieval*. Utrecht, Aug. 2007.

Notes

1. In this paper the same visualisation is used to highlight differences of "literal" citations in a historic context.