UNIVERSITÄT BONN

RHEINISCHE
FRIEDRICH-WILHELMS-
UNIVERSITÄT BONN

Institut für Germanistik der Uni Bonn · Am Hof 1d · 53113 Bonn

PhD and Academic Qualifications Office

Faculty of Arts, Charles University

Jan Palach Square 2

116 38 Prague 1

Czech Republic

Prof. Dr. Kristian Berg

Institut für Germanistik,
Vergleichende Literatur-
und Kulturwissenschaft

Germanistische Linguistik

Am Hof 1d
53113 Bonn
Tel. 0228/73-7426
kberg@uni-bonn.de

www.germanistik.uni-bonn.de

Bonn, 06.09.2021

External Examiner's Report on the Dissertation of Marie Vaňková,

**"Analysing Early Middle English online:
Construction and Use of a LAEME Based Spelling Database"**

Submitted in 2021 at the Department of English Language and ELT Methodology

I.
This dissertation develops new tools to explore and analyze the richly annotated existing data from the Linguistic Atlas of Early Middle English (LAEME). On the basis of existing research on Early Middle English (eME), particularly the thorny issue of reconstructing its phonological system, the author develops a) a database that makes only a minimum of presuppositions, and b) tools to query these data. She demonstrates the usefulness of both with several case studies.

II.
Overall, this is a good thesis which makes a valuable contribution to the field. The author develops new and innovative tools for the investigation of eME texts. These tools will – once they are freely accessible – prove to be a boon for researchers. Previously tedious, manual analyses can be generated with a few clicks, promising to yield new insights into the spatial and temporal development of eME. The use of technical terms (e.g. *litterae*, *words*) is not always consistent, though (see below).

III.
1. The dissertation is clearly structured, and the structure is reasonable. At almost every point in the text, the author guides the reader, often explicitly. Some more examples would have been helpful at times, though (e.g., p. 72). The author's aim is to develop a database and a tool to access existing data. To this end, she first reviews the relevant literature on historical dialectology as well as existing electronic sources (ch. 2). She competently discusses the notoriously difficult issue of sound reconstruction, and she compares the slightly differing concepts of *letter*, *grapheme*, and *littera* used by different authors. The learnings from this chapter are the basis for the presentation of the author's database and interface design (ch. 3), which is presented in a very reflected manner. The tools developed in this

UNIVERSITÄT BONN

RHEINISCHE
FRIEDRICH-WILHELMS-
UNIVERSITÄT

Institut für Germanistik,
Vergleichende Literatur- und
Kulturwissenschaft

chapter are then put to use (ch. 4), before the whole endeavour is critically assessed (ch. 5). Despite minor drawbacks, the author succeeds in developing a useable tool to easily access the existing LAEME data.

2. As for the formal aspects, the dissertation is inconspicuous as far as references, abbreviations, and footnotes are concerned; it does include a relatively large number of spelling errors, though (e.g. p. 23, *repreented > represented*; p. 43, *there is at attempt > there is an attempt*; p. 51, *Angush > Angus* etc.). I am attaching an annotated PDF version for the author in which I highlight some of the errors (I also mark some inconsistencies and ask some questions).

3. Methodologically, the thesis is very transparent. The author motivates each choice concerning the database structure or tool functionality, and she often reflects critically on her own achievements, pre-empting most of the critique. Having said (well, written, and this will become important in a moment) that, there are at least three potential problems I have with the argumentation.

Firstly, I find it curious that the author draws heavily on the work of devoted 'autonomists' like Vachek and MacIntosh, and then falls short of a purely graphemic implementation: The sets of litterae are defined by their supposed sound value (p. 71), and throughout the thesis, the letters' sound values serve as arguments for or against a specific analysis. It is only in the last part (166ff.) that the author shows what a purely graphemic analysis could really look like. It is very interesting that even if we restrict ourselves to variation within texts, differences between scribes become strikingly visible.

A second minor drawback is that the interpretation of the data is often somewhat shallow. A little more detailed and nuanced analyses would make an even better case for using the tool. For example, the conclusion on p. 134 ("There seem to be two basic kinds of sets, namely {ch, c, k} and {ch, h}.") does not follow from the analysis shown on the previous page.

And finally, there is a certain carelessness regarding theoretic fundamentals. For example, the term *littera* is borrowed from the existing literature. On p. 22f., the author states "*Letter* is practically coreferential with *littera* and *realisation* is a label for an individual instance of a letter". This distinction is certainly reasonable and is reminiscent of the term grapheme as a more abstract unit, encompassing letters (the letters *f*, *F* would be part of the grapheme *f*, for example) However, on p. 83, Figure 11, the author states that *w* is a littera with the 'characters' *p* and *f* (and by the way, why don't the ratios add up to 1?). This implies a much broader concept of abstractness, one that involves sound correspondences. Another instance is found on p. 62, where the author introduces *tags*: "Each tag consists of the actual *form* found in the MS, the so-called lexel, […] For example, the lexel AFTER […] has a wide range of forms, for instance aftir, hafter, eftre, affeter, hefteir etc." But that is again somewhat imprecise: Each word is *represented* by a tag; each tag *consists*

of the ACTUAL form "found in the MS" – but then the lexel "has a wide range of forms". How can that be if it *is* the actual form?

4. As mentioned above, the result of this thesis is a novel and original tool that can be used to access a richly annotated eME corpus. As such, it is a valuable contribution to the field which may lead to interesting generalizations about the spatial and/or temporal development of English dialects. Of particular relevance in this regard is the focus on visualizations, which allow the user to grasp basic characteristics of data sets immediately (cf. Beck & Butt 2020).

IV.
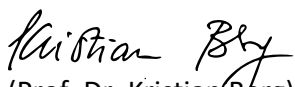I have the following questions for the author:

- **How does the proposed distinction between prodigal and economical** (alphabetic, it should be specified) **writing systems relate to Katz & Frost's (1992) idea of 'shallow' vs. 'deep' writing systems?**
- On p. 142, the author uses network visualizations to compare two manuscripts and concludes (on p. 143), "The relative economy of The Ormulum can be contrasted with the extreme level of prodigality found in the previously analysed MS Cambridge, Trinity College B14.39 (scribe A, #246)." **Could the author elaborate on the role of respective text length for the comparison?**
- A purely graphemic analysis (for which I expressed sympathies above) of course leads to alignment problems. Without reference to sounds, it is harder to align e.g. *knight*, *knit*, and *cnith*. **Do you think syllabic information (pertaining to the written syllable alone) could be a viable alternative?** We can determine consonant and vowel letters by distribution (Berg 2012), and also the ones in-between (e.g., *w*), and work from there. The advantage would be that we could more or less automatically group lexemes together, e.g. position 1 in *no*, *night*, *never*, something that is possible only manually in the tool now.

V.
I recommend the submitted dissertation with the tentative grade of

pass.

Bonn, 06.09.2021

(Prof. Dr. Kristian Berg)