A Metrical Analysis of Medieval German Poetry Using Supervised Learning

Christopher Hench* and Alex Estes

Department of German, University of California, Berkeley, Berkeley, CA, United States

Middle High German (MHG) epic poetry presents a unique solution to the linguistic changes underpinning the transition from classical Latin poetry, based on syllable length, into later vernacular rhythmic poetry, based on phonological stress. The predominating pattern in MHG verse is the alternation between stressed and unstressed syllables, but syllable length also plays a crucial role. There are a total of eight possible metrical values. Single or half mora syllables can carry any one of three types of stress, resulting in six combinations. The seventh value is a double mora, i.e., a long stressed syllable. The eighth value is an elided syllable. We construct a supervised Conditional Random Field (CRF) model to predict the metrical value of syllables, and subsequently investigate medieval German poets' use of semantic and sonorous emphasis through meter. The features used are: (1) the syllable's position within the line, (2) the syllable's length in characters, (3) the syllable's characters, (4) elision (last two characters of previous syllable and first two characters of focal syllable), (5) syllable weight, and (6) word boundaries. Additional metrical rules are enforced and marginal probabilities are calculated to yield the most likely legal scansion of a line. The model achieves a weighted average F-score of 0.925 on internal cross-validation and 0.909 on held-out testing data. We determine that trochaic alternation with a one syllable anacrusis and words carrying clear stress assignment are the easiest for the model to scan. Lines with multiple double morae of syllables with few characters are the most difficult. We then rank all the epic poetry in the Mittelhochdeutsche Begriffsdatenbank (MHDBDB) by the difficulty of the meter. Finally, we investigate the double mora, which MHG poets used to draw attention to chosen concepts. We conclude that poets generally chose to use the double mora to emphasize highly sonorant words.

Keywords: poetic meter, Middle High German, scansion, epic poetry, supervised machine learning, medieval literature

1. INTRODUCTION

Poetic meter in the Middle High German (MHG) tradition has always been a contentious and complex subject, as it requires a nuanced knowledge of MHG literature, a strong understanding of MHG linguistics, particularly phonology, and knowledge of the musical practices of the period[1]. Most work thus far has not been able to master all of these areas[2]. While this paper does not attempt to fully unite these diverse fields, it does seek to take careful consideration of each in developing a computational model to better understand how medieval German poets crafted their words into meter, and in turn aid us in our own reading of the text. The increased popularity of machine learning algorithms and their application to textual data presents a particularly fruitful opportunity in a domain that has plagued MHG scholarship for years. Instead of a deductive approach, i.e., beginning with the assumption of trochaic alternation as fundamental, supervised learning allows for a large-scale inductive approach, supplying the algorithm with a wealth of specific examples from which general principles can be discerned. Crucially, the goal of any such model is not to establish an absolute truth about a historical language; the goal is to automatically reproduce the annotation decisions of scholars on a large scale. Annotating the entire MHG epic corpus would allow us to better understand any rules that do exist as well as the challenges any particular text poses. Automatic annotation would also support a large scale analysis on how specific metrical values and meter types are invoked in different contexts. Scholars often discuss how changes in meter, metrical values, or specific cadences are triggered in specific scenes, but can we measure this complexity? MHG meter provides for fascinating flexibility in emphasis, but did authors have preferences for different metrical values? Are certain texts or passages intentionally crafted to be more difficult to scan? This paper seeks to answer these questions and others through a large scale analysis of automatically scanned poetry.

While late twentieth century scholarship neglected meter primarily due to theoretical disagreements and a lack of manuscript evidence, Christoph März recently re-framed MHG scholarship on meter in his article "Metrik, eine Wissenschaft zwischen Zählen und Schwärmen," in which he attempts to revive a meter-based formal perspective (März, 1999). According to März, form has two important functions and opportunities: it *reminds* us, and it allows for *comparison* (März, 1999, p. 325). Both of these observations provide motivation for the following analysis. Poetic meter acts not only to support the memory of a performer or composer, but also affects the audience, prompting this comparative reception. März writes:

I recall the experience that when you try to remember a poem, you often only remember the pattern—a few words may come along with that pattern or not. Also, if you forget parts of the text, the threads can be found again in certain passages by humming the rhythm of the verse (März, 1999, p. 325).

This act of remembering serves as an opportunity to identify connections between songs and texts (both formally and semantically), and compare texts, as März would have it. This comparison, when recognized by a performer or audience, can generate and add meaning to a poem or song. Especially in the MHG tradition, a connection between form and content has always been presumed. Yet März is also interested in lower level connections and references within genres. März asks whether these abstracted metrical schemata "transport" specific ideas, and if so, how they are created (März, 1999, p. 325). Klaus Kohrs asked a similar question decades earlier. Kohrs explains in Saussurian terms how meter itself can add signification to language, which it does not inherently carry: "With the metrical, that is even "quasi-musical" formation of language as a symbolic and sonoric phenomenon the side of the signifié is quasi sublimated, i.e., sensical and semantic references become virulent, which the "natural" language does not have and does not need to have" (Kohrs, 1969, p. 605). Hugo Kuhn presents the idea similarly in relation to music and melody, but emphasizes its "Gebrauchsfunktion" (use function), i.e., the use cases for these artworks, as folksongs, religious uses, for the court, knights, etc. (Kuhn, 1969, p. 38). This point is taken up by Thomas Cramer, questioning what the actual *Gebrauchsfunktion* for these artworks was, and whether our ideas of them are correct according to the sources (Kuhn, 1969, p. 39). But März crucially reshapes this question, instead of asking what meaning or function poetic meter may contain, he notes that meter is always determined relatively (März, 1999, p. 325). As Paul Zumthor and Ferdinand de Saussure have claimed about words and sound, there is no meaning in the base element itself, only in context and pattern. But for both words and poetic meter, this context must be extended beyond the contained object of a line of poetry to the body of referential objects.

The aim of this paper is to disambiguate these relative relationships. This project does not intend to argue that any particular metrical theory is without fault, nor that specific metrical values even exist as such, but rather that implementing *any* framework inevitably teases out relative differences within a corpus. Heusler decries the nineteenth century philologists for altering the text and making a statistical analysis of MHG metrics impossible, and for this reason he gives no statistics in his MHG study of meter (Heusler, 1956, p. 4). Yet the focus here is not concrete, in terms of absolute numbers or statistics, but rather in establishing relative relationships between texts, which, when aggregated over a whole text or corpus, will not drown out clear characteristics[3].

---

[1]This article has adapted much of the authors' previous work in "Supervised Machine Learning for Hybrid Meter" (Estes and Hench, 2016). Shortly after, Agirrezabal et al. undertook a similar project using the benchmark results set by Hench and Estes (Agirrezabal et al., 2016). It has also been cited by work adapting a scansion model to Spanish and Portuguese (Mittmann, 2016; Navarro, 2017).

[2]Christoph März criticizes recent scholarship as being too linguistic in nature, and forgetting the aesthetic effect meter produces (März, 1999, p. 323).

---

[3]Admittedly, idiosyncrasies in a specific text will, for this reason, be very difficult, if not impossible, to discern if edited texts are used for analysis.

2. FROM QUANTITY TO QUALITY

The distribution of Latin into distinct regional dialects had profound linguistic and literary implications for all of Europe. One notable consequence was on poetry with quantitative meter. Even before the Middle Ages, the syllable length of classical Latin had been nearly forgotten in the vernacular[4]. Latin poetry had used quantitative meter, in which syllable length is the organizing principle. Syllable length was a phonologically distinctive feature to Latin speakers. However, the emerging dialects differed from Latin in that stress became a phonologically important feature, and thus so-called qualitative meter ("rhythmic poetry") predominated in the Romance languages. Reconciling these linguistic differences, MHG meter relied on both stress and syllable length. This hybrid metrical form poses unique challenges to scanning poetry and allowed for a diverse development in genre and style (Heusler, 1956, pp. 74–75). Yet this freedom raises one of the main questions and theoretical problems in MHG research on meter: not necessarily Heusler's question of "*How* am I to measure it?," but rather *what*: in a system of "measured syllable verse with free syllable counts"—what is it that we *can* count, or *should* count? (Heusler, 1956, pp. 9, 13) "What is countable in the verse?" (März, 1999, pp. 323–324) We could count syllables, but it is not clear if the poets did this naturally themselves, despite what the *Meistersänger*[5] would like us to believe. Herbert Bögl describes MHG verse in his *Abriss der mittelhochdeutschen Metrik: mit einem Übungsteil*: MHG "presents in an abstract language of symbols the sequence of syllables in a verse and weighs them taking into account their length and stress" (Bögl, 2006, p. 9). It is this "taking into account" that presents a difficult computational problem for analysis MHG meter, in that strict rules for length and stress cannot always be employed.

To illustrate this shift from a quantitative classical meter to a qualitative post-classical vernacular meter, we first consider the quantitative epic poetry of Latin and Greek. Each line consists of six feet, each foot typically a dactyl (a long syllable followed by two short syllables) or spondee (two long syllables). A syllable is considered long if it has a long vowel or diphthong, or ends in two consonants (Hayes, 1989). All other syllables are short. The first line of Virgil's *Aeneid* serves as example:[6]

ārma vi|rumque ca|nō, Tro|jae quī|prīmus ab|ōrīs[7]

— ⌣⌣ | — ⌣ ⌣ |— —|— —| — ⌣ ⌣|— —

A widely cited poem displaying the shift from quantitative to qualitative rhythmic poetry in the Latin tradition is Bishop Auspicius of Toul's late fifth century letter to Arbogast, the Count of Trier, imitating the iambic dimeter[8] already made famous by Ambrose[9]. The letter begins:

Praecelso exspectabili   his Arbogasti comiti
Auspicius qui diligo   salutem dico plurimam[10].

The first hemistich[11] — ⌣ ⌣ — ⌣ — ⌣ ⌣ shows that a quantitative scansion would be ill-fitted to the rest of the verse, and that a strictly iambic scansion is preferred with a paroxytone[12] in the cadence. Much Latin poetry followed suit, and the medieval *Codex buranus* famously bears witness to the intermingling of Latin and MHG rhythmic verse, clearly demonstrating that they were drawing from the same rhythmical schemata. Germanic verse, on the other hand, did not originally follow the quantitative meter of antiquity, preferring organization according to alliteration and stress. In fact, Heusler calls alliteration the "Hausmarke" (house brand) of the Germanic language family (Heusler, 1956, pp. 92–93). In addition to alliteration, a further marker of Germanic verse is the *Langzeile* (long line), traditionally consisting of two *Kurzzeilen* (short lines), an *Anvers* (first half of the line) and *Abvers* (second half of the line) (Heusler, 1956, p. 100). While this tradition began earlier, a classic example of Germanic alliterative verse is the ninth century Bavarian *Muspilli*:

…sin tac piqueme,   daz er touuan scal.
uuanta sar so sih diu sela   in den sind arheuit,
enti si den lihhamun   likkan lazzit,
so quimit ein heri   fona himilzungalon,
daz andar fona pehhe: dar pagant siu umpi.
sorgen mac diu sela,   unzi diu suona arget,
za uuederemo herie   si gihalot uuerde[13].

This alliterative verse dominated throughout most of OHG and continued strong in the Nordic traditions. Around the same time that the *Muspilli* was written in the southeast, in the west Otfrid von Weißenburg in Alsace was beginning to incorporate characteristics of Old French poetry into his ninth century Old High German (OHG) verse. Otfrid's decision to incorporate end rhyme (referred to as a strictly Romance language influence by Heusler) is the first attested instance of Germanic poetry's break from the alliterative tradition. Thus Otfrid is generally considered the starting point for a study of modern German verse[14]. Otfrid's *Evangelienbuch* became the model for this new Germanic verse, though he retained the *Langzeile* from the older Germanic tradition. Otfrid established many of the new metrical possibilities in cadence (monosyllabic full, bisyllabic ringing, and

---

[4]Augustine writes toward the end of the fourth century that while he recognizes time intervals, he can no longer distinguish between long and short syllables: *syllabarum longarum et brevium cognicionem me non habere...* "I cannot recognize long and short syllables..." cf. Augustinus, *De musica*, III, 3, 5.

[5]"Master singers" of the following generation.

[6]— represents a long syllable and ⌣ a short syllable.

[7]"I sing of arms and of a man, he was the first [to journey] from the coasts of Troy..."

[8]Two feet per line of an unstressed syllable followed by a stressed syllable.

[9]See Norberg and Ziolkowski (2004, p. 101) and Mathisen (2003, p. 34).

[10]"The highest Count Abrogast waiting for Auspicius, whom I love and say many greetings." Text from Norberg and Ziolkowski (2004, p. 101).

[11]Half of a verse.

[12]Stress on the penultimate syllable.

[13]Text from Braune and Ebbinghaus (1994) "his day comes, on which he will die. If the soul then quickly makes it way and leaves the body lying there, then one army comes from the stars and another from hell: they fight over it (the soul). The soul may be worried until judgement is made as to which of the armies it will be brought."

[14]The break with alliteration was much stronger on the continent than in England, as demonstrated by Old English and Old Norse verse (Heusler, 1956, p. 8).

trisyllabic ringing) witnessed in the MHG period (Heusler, 1956, p. 13). Much of the influence on Otfrid's style came from various writings on religion, heroic stories, and charms recorded at the time[15]. Heusler argues that this freedom in verse came primarily from the church, specifically church songs. Heusler writes: "song more easily takes advantage of the prosodic freedom" (Heusler, 1956, p. 32). Concerning rhyme, for nearly 300 years there was only pair rhyme in the AABB form, occasionally AAA, until around 1150 (Heusler, 1956, p. 12). Otfrid's rhyme began as pure monosyllabic rhyme, and later developed into multi-syllable assonance and other types (Heusler, 1956, p. 20). As the importance of rhyme grew, it became necessary for the rhyming syllable to also carry accent (Heusler, 1956, p. 24,3 1). This new rhyme and accent provided an alternative means to tie verses together, but also ushered in new freedoms of measuring verse, as rhyme required syllables to relate to one another, something emphasized by the contemporary musicologists (Heusler, 1956, p. 9). The form of the Ambrosian hymn is the closest metrically to Otfrid. The greatest difference lies within the construction of the line, where the syllable count is not certain, and divided lifts[16] are abundant (Heusler, 1956, p. 35).

Otfrid's founding of the Germanic rhythmic verse was what Heusler calls a "Germanicizing" of the Romance iambic verse: free filling of verses with syllables, anacrusis, and more varied cadences (Heusler, 1956, p. 36). Heusler charts out the development of Germanic verse and its influence from the Romance tradition, particularly in that the mixing of alliterative and pair-rhyme verse led to the early Germanic free filling of feet. Yet metrical conventions did exist in Otfrid's verse. The last foot was still strictly monosyllabic and verses could range from four to ten syllables, but were more often somewhere in between (Heusler, 1956, p. 43). OHG verse often had feet with more syllables than MHG because OHG words simply had more medial syllables[17]. In contrast to MHG verse, OHG verse was more consistent with syllable length and duration (Heusler, 1956, p. 56). In this sense, OHG verse was a "mediator" between Latin and alliterative verse (Heusler, 1956, p. 63). With Otfrid, the German pair-rhyme *Vierheber* (four stresses per line) began to take shape:

Uuas | líuto | filu in | flí- | ze[18],
    ×| ×́ ×| ⌣ ⌣ ×| — | ×̀ ^

in | managemo | ága | lei | ze[19],
×| ⌣ ⌣ ⌣ ⌣ | ×́ ×| — | ×̀ ^

3. THE MIDDLE HIGH GERMAN *VIERHEBER*

The most comprehensive and still referenced study of German meter is Andreas Heusler's three volume *Deutsche Versgeschichte*.

Heusler's theory has been criticized incessantly over the years, but persists as the accepted theory for MHG meter today. März claims that as reluctant as we are to use Heusler's theories, we use them because there is simply no better alternative (März, 1999, p. 318). While attempts have been made to supplement or critique Heusler's work, especially the existence of the fundamental "Takt" (measure, as in music), it has proven difficult for alternative theories to escape temporal restraints. If there is no "Takt," is there no foot, or stress alternation? (März, 1999, p. 319) As März observes, many of the alternative theories do not differ significantly from Heusler's, only Franz Saran's "*Schallanalyse*" (acoustic analysis) is suggested by März as a plausible alternative to better incorporate the actual voice of the verse (März, 1999, pp. 321–322).

What follows is a description of MHG epic meter in the Heusler tradition, with supplement from other, mostly pedagogical, resources. The Heusler theoretical framework is then employed to construct a supervised machine learning model of scansion[20].

The predominating pattern in all MHG verse is an alternation between stressed and unstressed syllables (Tervooren, 1979). MHG epic verse employs trochaic tetrameter: each line has four feet, and each foot is a trochee; this is known as the *Vierheber* for the four lifts (stressed syllables) in a line. Phonologically, a trochee consists of two syllables; the first syllable is stressed, and the second is unstressed. For example, the English word "better" is a trochee, but the word "alive" is not. The famous Longfellow epic poem *The Song of Hiawatha* is written in trochaic tetrameter, and the first line serves to illustrate this rhythm:

Should you |ask me, |whence these |stories?[21]
| ×́   ×| ×́ × | ×́   ×| ×́ ×|

Similarly, the prototypical MHG epic verse foot is two syllables in length, a stressed syllable followed by an unstressed syllable. However, feet can also be filled by one or three syllables (Domanowski et al., 2009). If a foot is filled by one syllable, the syllable must be phonologically heavy (containing a long vowel or ending in a consonant). If the foot is filled by three syllables, either the first two or the last two syllables are often phonologically light[22].

It is in these atypical feet that the influence of quantitative meter, where syllable length is a key factor, becomes evident in MHG verse. The foot in a *Vierheber* must be slightly redefined to account for this. Phonologically, syllable length is measured in morae, a unit of time such that a short syllable has one

---

[15]Incantations saw the greatest innovation in a move from older forms of verse to a distinctly Germanic verse (Heusler, 1956, p. 6).
[16]Akin to eighth notes in music.
[17]See Heusler (1956, pp. 48, 126).
[18]Braune and Ebbinghaus (1994). "There were many hardworking peoples,"
[19]Braune and Ebbinghaus (1994). "with such great zeal,"

---

[20]While Heusler's theory is certainly debatable, it must be reiterated that the intent of this work is not necessarily the absolute, *real* meter of MHG, but once again *relative* differences throughout the corpus, which could be revealed by an array of different theories. Furthermore, Heusler does not note any serious differences in meter between regions of MHG, from Heinrich von Veldeke in the north, to the Austrian southeast, to the Frankish northwest, there was little variation in the general *Vierheber* (four stresses per line) patterning (Heusler, 1956, p. 77).
[21]Longfellow (1932).
[22]Excepted are several end syllables in divided falls such as "-er," "-el," and "ez" (more below) (Domanowski et al., 2009).

mora and a long syllable has two morae (Fox, 2000)[23]. A foot in MHG meter is more precisely defined as having two morae, not necessarily two syllables[24]. Indeed the mora, not the syllable, has been called the fundamental unit of MHG verse, although the mora functions differently in this poetic tradition than in its phonological definition (Tervooren, 1979, p. 1). If a foot has only one syllable, the syllable must be heavy because a heavy syllable is two morae and the MHG foot requires two morae. A light syllable cannot be the only syllable in a foot, since it cannot be two morae. If a foot has three syllables, two are often light because half morae are most often light syllables (the first half mora of a pair must always be light), together forming one mora[25]. The other syllable is analyzed as one mora, yielding the required two morae in the foot. To summarize, a syllable can have one of three length values: mora, half mora, or double mora. A half mora must be phonologically light, and a double mora must be phonologically heavy. Phonological length is otherwise irrelevant and any syllable can be one mora (Heusler, 1956, p. 111).

In addition to length, as a function of morae, syllables are also assigned stress. There are three stress values: primary, secondary, and unstressed. Primary stress is assigned to the first or only stressed syllable in a word. Secondary stress is assigned to any following stressed syllable(s) in that word. All other syllables are unstressed[26].

The final mora of the final foot of a line is omitted by convention[27]. This is construed as a pause, analogous to a rest in music, and receives its own symbol in the scansion ∧, even though there is no corresponding word or syllable[28]. A short, word-final syllable may also be elided before a word beginning with a vowel. Finally, MHG epic verse permits up to three syllables in anacrusis (or pickup notes, a series of syllables at the beginning of a line that do not count in the meter). Phonologically, these syllables may or may not carry lexical or syntactic stress, but metrically, they are always scanned as unstressed morae.

The above features yield eight possible metrical values for any syllable:

1. **mora - primary stress** (x́): a syllable with primary stress
2. **mora - secondary stress** (x̀): a syllable with secondary stress
3. **mora - unstressed** (×): an unstressed syllable
4. **half mora - primary stress** (◡́): a short syllable with primary stress; according to metrical convention the preceding syllable must be long (Tervooren, 1979, p. 5)
5. **half mora - secondary stress** (◡̀): a short syllable with secondary stress
6. **half mora - unstressed** (◡): an unstressed syllable
7. **double mora** (—): a stressed long syllable; double morae always carry primary stress
8. **elision** (e): an elided syllable

Line 1 of Hartmann von Aue's *Der arme Heinrich* is prototypical. Each foot consists of a stressed syllable followed by an unstressed syllable. There is a one-syllable anacrusis:[29]

Ein | ritter | sô ge|lêret | was[30]
× | x́ × | x́ ×|x́ × | x́ ∧

Line 6 also begins with one syllable in anacrusis. The second foot has a stressed mora consisting of two syllables, each one a half mora. The third foot has one syllable; a diphthong allows it to be scanned as long. The final foot has a mora with secondary stress, since the preceding syllable is stressed and in the same word:

der |nam im |manege |schou |we[31]
× | x́ × | ◡́ ◡ × | — | x̀ ∧

Line 34 has no anacrusis, and in the second foot two half mora syllables form the unstressed mora:

|die ein |ritter in |sîner |jugent[32]
| x́ × |x́ ◡ ◡ |x́ × |◡́ ◡ ∧

Line 8 shows an elided syllable in the second foot:

dar |an be|gunde̦ er |suo|chen[33]
× |x́ × | x́ × | — | x̀ ∧

4. COMPUTATIONAL APPROACHES TO METER

There are two prevailing treatments of meter in the literature concerned with computational poetic text analysis. One approach takes a known meter and assigns syllables to stress patterns based on such parameters (Hartman, 1996). The second approach assumes nothing of the meter, and seeks to determine it by marking syllables and identifying patterns (Plamondon, 2006; McAleese, 2007; Greene et al., 2010; Agirrezabal et al., 2013; Navarro, 2015). This project draws more on the latter. Previous scholarship has also focused on relatively simple systems of meter and adopted rule-based, statistical, or unsupervised approaches. The hybrid nature of MHG meter, and other complex systems

---

[23]For example, the English word "red" has two morae since it ends in a consonant, whereas the first syllable in the English word "reduce" has one mora, since it ends in a short vowel.

[24]It can be helpful to think of MHG meter in the musical sense. Each foot is a measure of 2/4 meter, where one mora is equivalent to one quarter note, a double mora is a half note, and a half mora is an eighth note (Bögl, 2006).

[25]Occasionally very weakly stressed long syllables can also count as a half mora.

[26]The metrical distinction between different degrees of stress is rooted in phonological reality (Giegerich, 1985): in a word with many syllables, one syllable usually has a primary stress, and the others have either secondary or weak stress. For example, many pronounce the English word "anecdotal" with secondary stress on the first syllable, primary stress on the third syllable, and weakest stress on the second and fourth syllables.

[27]There are few exceptions to this in epic meter. The lyrical poetry, however, does break from this convention.

[28]The lyric tradition did allow for a feminine full cadence, which filled the entire last foot. This generally does not appear in the epic tradition.

[29]Text from Hartmann and Mertens (2005). Note that this notation differs slightly from that which is used for classical verse.

[30]"There was a knight so learned"

[31]"He looked extensively,"

[32]"Which a knight [should have] in his youth."

[33]"in [these books] he began to search,"

developing out of classical antiquity, makes it difficult to scan poetry using these methodologies, and thus supervised learning presents itself as an attractive method. After initial results of this project were published in Estes and Hench (2016), similar studies were undertaken for English in Agirrezabal et al. (2016), Spanish in Navarro (2017), and Portuguese in Mittmann (2016) with the results here serving as the benchmark.

4.1. Rule-Based Approach to Middle High German

A strictly rule-based approach to scanning MHG epic meter was undertaken by Friedrich Dimpel in 2004 (Dimpel, 2004a). As Dimpel's work is the only of its kind in this field, it deserves special consideration here. As part of his dissertation and continuing work at the University of Erlangen, Dimpel developed a set of tools named *ErMaStat* (*Erlanger-Mittelalter-Statistik*), crafted specifically for MHG epic poetry (Dimpel, 2004b). Although sure to admit the shortcomings of such an approach, the opening pages of his introduction to *ErMaStat* reveal his stylometric intentions in making such a suite of tools:

> Whenever one attempts to approach literary, scholarly questions with quantitative processes, then one must assume that texts from different authors (or different periods of an author's work) demonstrate certain distinct characteristics on a phonological, morphological, lexical, and syntactical level, which allow themselves to be captured quantitatively (Dimpel, 2004b).

Dimpel's list of variables include: (1) syllable, word, and line count, (2) vowel and consonant counts, (3) function words (specific parts of speech), (4) alliteration, assonance, and enjambment, (5) suffixes, (6) word frequencies, (7) prefixes, (8) common words (a finer measurement than word frequency), (9) word combinations (naïve bigrams), and (10) a metrical analysis. His intention is to model style, or characteristics of style, in order to compare texts and estimate probabilities of works being written by the same author.

Dimpel continues with three examples. In the first example, he takes four of the better known MHG epics: *Parzival*, *Tristan*, *Wigalois*, and *Willehalm*. Using the variables above, he calculates and averages significance values, showing that *Parzival* and *Willehalm*, both written by Wolfram von Eschenbach, do in fact have a lower degree of quantified stylistic difference relative to one another than to the works by other authors. Dimpel is also able to determine the contributions from individual variables. Dimpel's second analysis concerns the grouping of Wolfram's *Parzival* into chapters and the thesis proposed by Elisabeth Karg-Gasterstädt of four different sound types, following the work of Eduard Sievers (Karg-Gasterstädt, 1925). Dimpel's *ErMaStat* supports Karg-Gasterstädt's hypothesis as a possibility. His last example considers the date of authorship of Hartmann von Aue's *Iwein* with respect to Hartmann's *Erec*.

Dimpel approaches MHG meter by first programming for alternation and then hierarchically creating rules to account for stress. Though his work must be commended for its accuracy and linguistic engagement, it is a laborious task, inflexible, and extremely language specific. Our intention here is not to duplicate

his work, nor dismiss it. Rather, through supervised learning we offer a new approach to an old problem for MHG. It also provides an opportunity for the "drei-stufige" (three-level, i.e., accounting for secondary stress and double morae) scansion Dimpel has not yet attempted, but notes is a great challenge to modeling MHG meter. There are also advantages of particular interest to humanists. A supervised method will learn to scan more in the manner of a human than a strictly rule-based approach would, perhaps remaining truer to the poetic tradition, and giving insight into what poses difficulties for human scanners. It also allows for greater versatility, and a chance to analyze the prosody beyond the epic meter, and perhaps even prose (Dimpel, 2004b, 2015).

4.2. Supervised Learning

The machine learning approach taken in this project is supervised, i.e., the computer is provided with annotated data in the form:

ein/MORA   WBY/WBY   rit/MORA_HAUPT   ter/MORA
WBY/WBY sô/MORA_HAUPT WBY/WBY ge/MORA lê/
MORA_HAUPT ret/MORA WBY/WBY was/MORA_HAUPT

The algorithm then learns which of the annotated features (described below) are important, and subsequently how to classify any given syllable. In contrast to other automated scansion systems, a supervised approach learns how the human annotators scanned based on a set of provided features and annotated data, as the algorithm identifies which features were deemed important by the humans who annotate them. When working with human productions, such as poetry, this is an attractive advantage. Yet there are both advantages and disadvantages to this method. On the one hand, the resulting model will take contextual and situational factors into account, factors that a strictly rule-based approach may not, due to the multiple layers of rules and probabilities constructed. If a poet attempts a certain stylistic move during a section of the narrative it may be captured (via a combination of feature weights) by the model. On the other hand, because the model learns to scan poetry like the annotators, it necessarily follows the practices of a specific theoretical school, thereby narrowing the limits of interpretation.

5. DATA

Because supervised learning is a novel approach to poetic meter, annotated metrical data do not exist for MHG or most other languages. Following the Heusler scansion system outlined above, syllables of MHG epic poetry were annotated into the eight categories of metrical value. The annotated data consist of 450 lines from Hartmann von Aue's *Der arme Heinrich*, 200 lines from Wolfram von Eschenbach's *Parzival*, and 100 lines from Wirnt von Grafenberg's *Wigalois*[34]. An additional 10% (75 lines of Hartmann von Aue's *Iwein*) was annotated to be held-out for

---

[34]Incorporating different poems from different poets accommodates varying styles of writing, but it also introduces more variability, which will become the foundation of the model.

TABLE 1 | Summary statistics for annotated dataset.

	Mean	Std.	Min.	Max.
Char. per line	21.34	3.39	9	32
Syll. per line.	7.62	1.04	5	11
Words per line	5.01	1.13	1	8
Char. per word	4.26	1.96	1	17
Syll. per word	1.52	0.71	1	7
Char. per syll.	2.80	0.81	1	7

TABLE 2 | Inter-annotator agreement confusion matrix.

		Annotator 2						
		×	×́	×̀	—	◡	◡́	ẹ	◡̀
Annotator 1	×	285	4	0	1	3	0	0	0
	×́	0	225	1	0	1	0	0	0
	×̀	1	2	74	0	2	0	0	0
	—	1	2	0	72	0	0	0	0
	◡	1	0	0	0	36	0	0	0
	◡́	0	1	0	0	0	17	0	0
	ẹ	0	0	0	0	0	0	9	0
	◡̀	0	0	0	0	0	0	0	1

testing, yielding a total of 825 annotated lines. Summary statistics for all annotated data are reproduced in **Table 1**.

Syllabification was performed prior to annotation according to the system detailed in Hench (2017), i.e., following the Sonority Sequencing Principle (SSP) with correction from the Legality Principle (LP). Annotation was carried out by the authors, who are both trained in MHG scansion[35]. In the case that a line exhibits multiple permissible scansions, priority is given to the scansion that best preserves the alternation of stressed and unstressed syllables. If a decision still cannot be made, then stress is determined according to semantic importance. An additional consideration is the syntactic stress of a particular line. Clearly, such evaluations allow some room for interpretation. Nevertheless, on a sample of 100 lines from the annotated data (739 syllables), the Cohen's kappa coefficient for the inter-annotator agreement is 0.962 (confusion matrix given in **Table 2**). The greatest disagreement for the human annotators was among unstressed and stressed morae, and between unstressed morae and unstressed half morae, implying both some stress and some value disagreement.

6. WORKFLOW

We present a new workflow, illustrated in **Figure 1**, for the automated scansion of poetic meter (MHG and other). The

process begins with the syllabification of texts[36]. Because the syllable is the base unit for many poetic traditions, it is what needs to be annotated. After syllabification is the metrical annotation, requiring experts in scansion and the texts themselves[37]. After annotation, features must be developed and extracted to help the model understand what part of a verse is important in assigning metrical values. Many of the most important features are phonological. After these features are identified, an algorithm must be developed for obtaining and annotating these features before they, along with the syllables themselves, are sent to the model. After feature identification and extraction, there is a process of development and validating the model, in order to choose the most suitable parameters for the task. The model then makes predictions for each syllable based on the features and parameters supplied. The model also yields marginal probabilities for all predictions for any given syllable, i.e., each syllable comes with a list of probabilities for each of the possible metrical values. Simply taking the most probable sequence for any given line can be very accurate[38].

For epic poetry, and the model described here, the predictions must be further processed through specific constraints, as the model will not impose strict rules unless instructed to do so[39]. If the predicted sequence passes each constraint, it is considered the final scansion. If not, the line is sent for further processing. The following tests check for impossible scansions in MHG epic meter:

- **Four stresses**. The heart of the *Vierheber* is exactly four stresses per line[40]. Any more or less fails the line.
- **Double morae must be heavy**. Phonologically, to carry the weight of two morae, the syllable *must be heavy*. If a light double mora is identified, the line fails.
- **First syllables of divided lifts must be light**. While divided falls are allowed to be either light or heavy depending on the end syllable (though usually light), the first syllable of a divided lift *must be light*, if not, the line fails (Domanowski et al., 2009).
- **Elided syllables must be light**. An elided syllable must end in a short vowel.
- **Alternation**. Two lifts cannot follow one another unless a lift follows a double mora, otherwise stress alternation is not upheld and the line fails.

---

[35] Although neither author is a native speaker of NHG, the two phases of the language and the metrical traditions are sufficiently different that both native and non-native speakers require training in MHG scansion.

[36] Syllabification was performed following the method introduced in Hench (2017), which established an accuracy of 99.4% on MHG.

[37] For the presented model, the accuracy will be highest if the text is standardized and includes markers of long vowels because the annotated texts were such, and the extracted features depend upon this. However, the model can still scan any sort of text input with a reduced accuracy.

[38] Considering that there is not yet a constraint to the four stresses in MHG epic meter built in, this preliminary model without additional rules would be very helpful in eventually constructing a model for MHG lyrical poetry, which does not adhere strictly to the *Vierheber* qualities. In a sense, the bare model is a model best suited to predicting stress.

[39] Results for the model without constraints are an F-score of 0.894 on cross-validated data and 0.904 on held-out data (Estes and Hench, 2016).

[40] Generally, this rules out the *stumpf* (blunt) cadence, which carries only three stresses. The *stumpf* cadence is rarely the only possible scansion (often a double mora can be assigned to fill the feet), though as Bögl points out, some lines in *Erec*, for example, leave the *stumpf* cadence as the only possible scansion (Bögl, 2006, p. 26).

FIGURE 1 | Flow chart for MHG scansion.
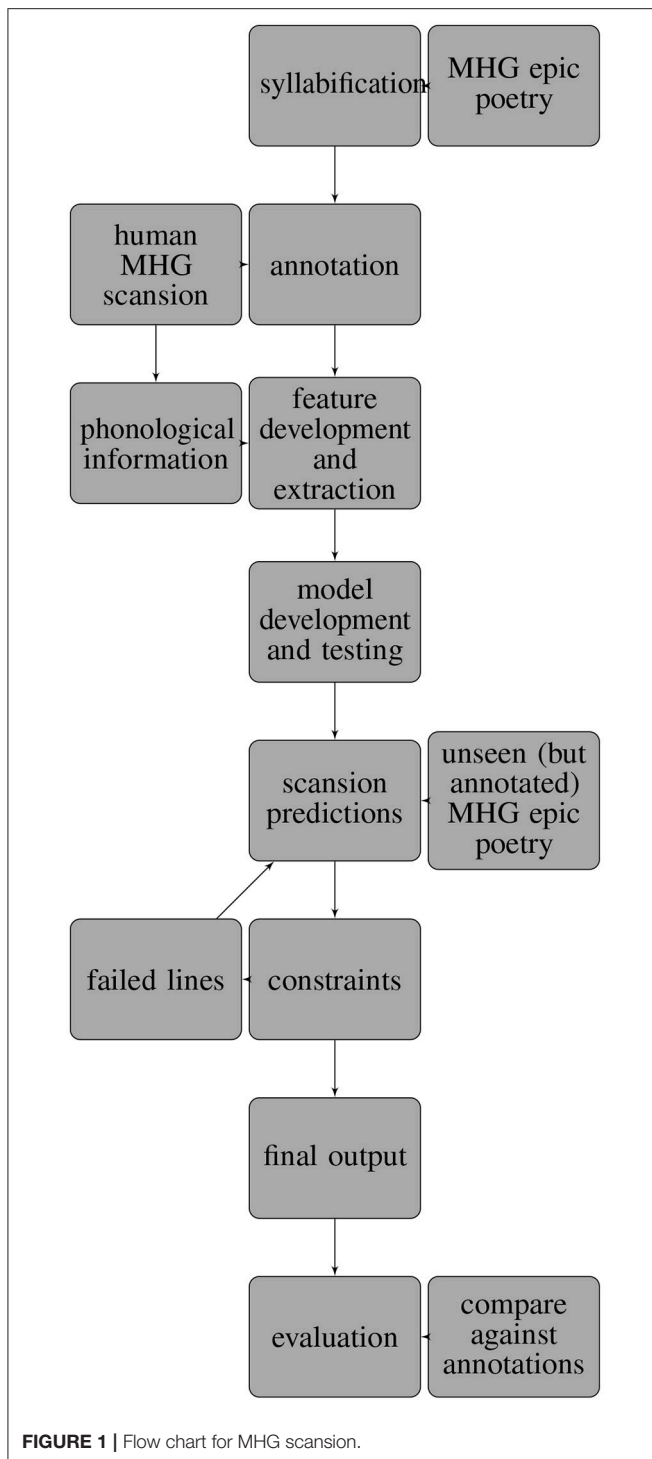
If the line fails any of the above tests, it is sent for further processing, where, depending on the phonology of the syllables and the rules above, all phonologically possible metrical values as well as the calculated marginal probabilities are combined into every possible sequence for all the syllables in the line. Illegal combinations according to the tests above are ruled out of the set, and the set is then ranked by overall probability. The

most probable, legal scansion is then selected. Thus, errors in the model stem primarily from incorrect identification of the language's natural stress, as all certain metrical patterning errors are sorted out. The model itself is intended to account for this natural stress by taking advantage of the information provided to it through the annotated data. The constraints help the model further cut out possibilities that conform to the natural stress, but not the metrical environment (which is relatively weakly learned by the model), yielding an ultimately highly accurate model.

7. MODELS AND FEATURES

Two baseline models were developed[41]: an n-gram model[42] cascading into regular expressions and a Brill transformation-based model on top of the n-gram model, both using syllables as units, though not explicitly accepting features beyond the syllables and tags themselves. The n-gram model consists of cascading trigram, bigram, unigram, and regular expressions models, i.e., first a value is predicted based on the previous two values, if possible; otherwise it is predicted based on the previous one value, and if the first two models fail, it is predicted solely based on the value probability for the syllable itself. If the syllable did not appear in the training data, and it cannot be predicted by the first three models, it resorts to regular expressions. Based on MHG scansion theory and observations while annotating, syllables with long vowels were assigned to double mora, short syllables to unstressed mora, and the remaining syllables to mora with primary stress[43]. The n-gram model was implemented with default settings, and is a very naïve most-frequent-tag approach. This model helps to illustrate the variation between appropriate metrical values for the same syllable.

The Brill model[44], first assigns the most common label for a given syllable and previous syllable's metrical value from the n-gram model described above and then generates rules to improve the initial estimate of the n-gram model according to the training data. It then iterates over these rules, correcting labels until F-score no longer increases. The Brill model was implemented with a maximum of 200 rules. This approach is very similar to Dimpel's enumeration of hierarchical rules for MHG scansion.

To compare to the baseline, efforts were focused on constructing a Conditional Random Field (CRF) model[45]. The decision to implement a CRF model was predicated on the interpretability of CRF modeling and understanding the primary features for MHG scansion[46]. In the model, each syllable contains

---

[41] The results for all models were internally 10-fold cross-validated and tested on held-out data. All code is available at https://github.com/henchc/FDH-2018

[42] Implemented with the help of the NLTK n-gram tagger, Chapter 5 (Bird et al., 2009).

[43] This proved important to recognize stress alternation.

[44] See Brill (1995); implemented with the help of NLTK (Bird et al., 2009).

[45] See Lafferty et al. (2001); The implementation of the CRF model was expedited with the help of *crfsuite* (Okazaki, 2007).

[46] A CRF model fits the problem of scansion better than a traditional Hidden Markov Model (HMM) because HMMs only consider relationships between each prior state and the observation. A CRF model relaxes the independence assumption and considers both previous and following labels in determining the label for a given syllable. This is especially helpful for scanning MHG meter. For

the features for the syllable itself, but also those of every syllable in the line, marked by index. The features and their motivations are:

- **Position within line**: the last mora of a line is always stressed (except in masculine bisyllabic cadences), and double morae occur most often in the third foot. If there is anacrusis, these syllables will be unstressed morae.
- **Length of syllable in characters**: longer syllables (in terms of number of characters, serving as a proxy for phonemes) are more likely to be stressed. Unstressed prefixes and suffixes tend to be maximally three characters.
- **Syllable characters**: the characters in a syllable can help identify certain grammatical morphemes that are often unstressed. Slices were taken of the first character, first two characters, last character, and last two characters.
- **Elision**: the last two characters of the previous syllable and the first two characters of the current syllable are identified as one feature to detect conditions for elision.
- **Syllable weight and length**: syllables ending in a vowel or consonant are open or closed respectively. Syllables ending in a short vowel are short; otherwise they are long. Such values are useful in identifying double or half mora syllables, which must be long or short respectively. For example, the syllable "schou" in line 6 of *Der Arme Heinrich* above is a double mora, and is accordingly long.
- **Word boundaries**: MHG is a stress initial language.

The model was tuned only on the cross-validated development data and the best performing model was chosen. The resulting best model uses an *L1* coefficient of 1.3 and *L2* coefficient of 0.001. No further changes to the model itself were made after the model features and parameters were selected[47]. However, the additional rules described above were enforced in order to increase the F-score for epic meter specifically.

8. RESULTS

The n-gram model found little success even with additional training data, ending with an F-score of only 0.602 (95% CI [0.583, 0.623]). The transformation-based Brill model improved quickly upon the n-gram model, but plateaued at an F-score of 0.810 (95% CI [0.796, 0.838]). **Figure 2** shows the increase in F-score with an increase in the number of annotated lines for all models, suggesting that marginal returns to annotation begin to diminish significantly after around 400 lines, or, in the case of MHG, about 3,000 syllables. The final results of the CRF model are given in **Table 3** in descending order of frequency in the data, along with a final held-out test set of 75 lines from Hartmann von Aue's *Iwein*. The preferred CRF model achieves an F-score of
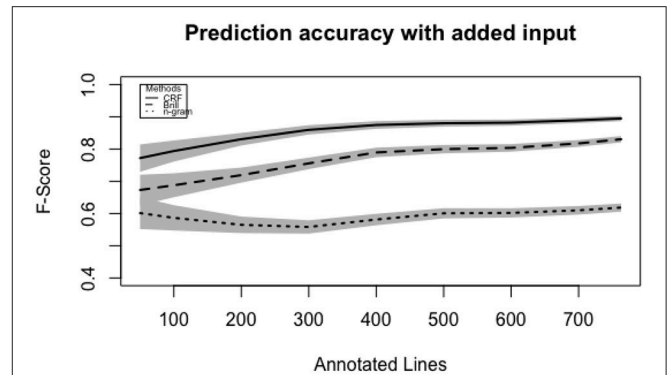

FIGURE 2 | Tagging F-score with added input.

TABLE 3 | CRF model F-score for individual metrical values and (weighted) average in development and on held-out data.

Metrical value	*F*	Obs.	Held-out *F*	Held-out Obs.
mora - unstr.	0.938	2,405	0.937	253
mora - prim.*	0.949	2,463	0.951	253
double mora	0.881	424	0.928	34
half mora - unstr.	0.672	231	0.541	41
half mora - prim.	0.822	107	0.667	11
elision	0.773	65	0.667	2
(Weighted) average	0.925		0.909	

Morae and half morae with secondary stress were not predicted, rather determined based on word boundaries after prediction.

0.925 (95% CI [0.911, 0.939]) on the cross-validated development data and 0.909 on the held-out testing data[48]. Supervised learning thus proves to also be an economical option for languages with complex meter.

The top ten highest scoring features of the CRF model and rules of the Brill model are given in **Table 4**. It is evident that the CRF model takes advantage of the phonological features provided, an advantage over the baseline models. Top CRF features (1) and (5) suggest any heavy syllable is likely stressed, and often a stressed mora. The CRF model also discerned cadence from the line patterning, exhibited in top CRF features (2) and (4), noting that except for the rather uncommon occurence of a cadence with a divided lift in the last foot (masculine bisyllabic cadence), the last syllable is always a stressed mora. Elision appears frequently in the top CRF features (3) and (6). Anacrusis is recognized in top CRF features (7) following the prototypical patterning:

Ein | ritter | sô ge|lêret | was[49]

× | x́ × | x́ ×|x́ × | x́ ^

---

example, if the model can see that there are likely two half morae at the end of a foot, then the beginning of the foot is likely one mora and not a double mora. Future work might consider an alternative in sacrificing interpretability for accuracy utilizing a Bidirectional Long Short Term Memory (BLSTM) neural network, though this is not attempted in this project because the CRF model proves very accurate when considering Cohen's Kappa.

[47] All features are encoded as categorical. The *crfsuite* package binarizes categorical variables.

[48] For reference, on the annotated data the model has a Cohen's Kappa of 0.949 with Annotator 1 and 0.944 with Annotator 2.

[49] "There was a knight so learned" (Hartmann and Mertens, 2005, p. l. 1).

TABLE 4 | Top ten CRF features and Brill rules.

CRF	Brill
(1) — if heavy syll.	(1) x́ → × if at word boundary and following syll. is x̀
(2) not x́ if next syll. is EOL	(2) x́ → — if followed by x̀ and word boundary
(3) ẹ if first char. of next syll. is "e"	(3) × → x̀ if end of line
(4) x́ if EOL	(4) x̀ → x́ if monosyllabic
(5) x́ if heavy syll.	(5) × → x́ if following syll. is "ge"
(6) ẹ if first char. of next syll. is "i"	(6) x́ → × if "ist" +2 syll.
(7) × if +7 syll. not EOL	(7) ẹ → ◡ if "der," "den," "diu" +2 syll.
(8) — if next syll. same word	(8) ◡ → ẹ if "ein," "ich," "er" +2 syll.
(9) — if next syll. ends in "en"	(9) ẹ → x̀ if "ge" +2 syll.
(10) not ◡ if beg. of word	(10) — → × if "al" -1 syll.

TABLE 5 | CRF confusion matrix.

		Predicted					
		×	x́	◡	—	◡̆	ẹ
True	×	239	4	8	2	0	0
	x́	6	243	3	0	1	0
	◡	11	7	20	1	0	2
	—	0	1	1	32	0	0
	◡̆	1	3	1	0	6	0
	ẹ	0	0	0	0	0	2

Where there are eight syllables, if seven syllables down the line is the last syllable of the line (EOL), and alternation is regular, that focal syllable will be unstressed in anacrusis. Top CRF features (8), (9), and (10) each consider words and word boundaries, specifically that double morae often occur at the beginning of a multi-syllabic word, and often that word is bisyllabic, with the second syllable ending in "en," such as "mae-ren" or "rî-ten." Top CRF feature (10) notes that unstressed half morae often occur after the first syllable of multi-syllabic words (half morae are only stressed when beginning a word).

The Brill model adopts a more general rule for alternation in top Brill feature (1). Notably, the Brill model takes greater advantage of word boundaries in (1) and (2), while these features rank lower in the CRF model. The Brill model inevitably also notes the influence of specific words or prefixes. The unstressed prefix "ge" ranks as a top five rule for the Brill model. If "ist" is two syllables down the line, the current syllable's assignment is changed from stressed to unstressed, or if "ein," "ich," or "er" is two syllables away, the original half mora assignment is changed to an elision.

The scores from both models confirm extant MHG metrical theory (as it was employed for the annotation), but suggest new methods of approach for students of MHG meter. Instead of first marking stress, as suggested by *Minimalmetrik* (Tervooren, 1979) and the pedagogically oriented website *Mittelhochdeutsche Metrik Online* (Domanowski et al., 2009), it may be useful for students to first determine the cadence and anacrusis by counting the number of syllables in the line, and looking for heavy syllables at the end of the line. Stress can then be marked in the remaining syllables and metrical values can be assigned based on phonological features. This method may be particularly useful for non-native German speakers, who may have less feeling for the natural stress of German. These results and insights support our feature decisions and our implementation of a CRF model.

9. ERRORS AND CHALLENGES

Investigating the errors and challenges of a supervised model presents the opportunity for the greatest new insights into the field and the advantage over an unsupervised approach. The confusion matrix for the CRF model in **Table 5** shows the errors made in the prediction of the held-out data. The model has the most trouble predicting both stressed and unstressed half morae, particularly the latter. This situation is mirrored in the human inter-annotator agreement matrix, demonstrating, as may be expected, that the machine learning model makes similar errors to the human annotators. The unstressed mora and half mora confusion, common in both human and machine annotation, is understandable, as these two are the most phonologically ambiguous metrical values in MHG meter. Double morae, stressed half morae, and elisions all have the phonological restrictions listed above, and stressed morae are evidently less confused with double morae, likely due to clear alternation in the surrounding environment. Unstressed morae and half morae have hardly any restrictions (only that they are likely not heavy syllables). This then generates further stress confusion between unstressed morae and stressed morae seen in both the computer model and human annotation.

If an algorithm can be trained to scan MHG meter similar to how a human does, it may be interesting to see what is considered difficult for the algorithm. This can be computed by averaging the marginal probabilities calculated by the CRF model. The lower the average probability for a line, the less confident the model is about its provided scansion, and vice versa. With the model, this can be computed for any text (annotated or not), but let us first look at the model text used extensively for annotation and instruction, Hartmann's *Der arme Heinrich*. Unsurprisingly, the easiest lines for the model to scan are lines that hold true to the trochaic tetrameter patterning, inclusive of the common one syllable anacrusis:

(1) sus | trouc ouch | mich mîn | tumber | wân[50]

   × | x́   × | x́   × | x́   × | x́ ^

(2) ge | frumten | sô ge| sundez | hin[51]

   × | x́   × | x́ ×|x́   × | x́ ^

---

[50]"Thus I was also deceived by my foolish belief," l. 400. Average marginal probability: 0.9994.

[51]"Beneficially so healthy (their child) in (to death)" l. 1034. Average marginal probability: 0.9991.

(3) Ein | ritter | sô ge|lêret | was[52]

×͏ | ´× × | ´× ×|´× × | ´× ^

(4) ich | weiz wol | daz er | selbe | giht[53]

× | ´× × | ´× ×| ´× ×| ´× ^

(5) er | ist ein | vil ver| schaffen | gouch[54]

× | ´× × | ´× ×| ´× × | ´× ^

In these examples we also see very distinct prosodic stress patterning. "sus," "ouch," "mîn," "ber," "ge," "ten," "ge," "dez," "ein," "ter," "ge," "ret" are all common unstressed MHG syllabes both in poetry and prose, while "trouc," "mich," "tum," "wân," "frum," "sô," "sun," "hin," "rit," "sô," "lêr," "was" are all either semantically significant, or heavy, stressed syllables. Moreover, the stress of all multi-syllabic words is clear: "gefrumten," "gesundez," "gelêret," and "verschaffen" have unstressed prefixes, while the rest follow the typical word-initial stress. Any MHG scholar would notice the clear trochaic quality of these lines, and few would disagree with the typical trochaic scansion. One must also wonder how the poets understood and wrote these lines. Were they particularly easy to craft? Do they carry less importance in the story? Or are these lines particularly *true* (and just not informative?), because the poet did not need extra effort to fit the truth into a legal line of MHG poetry? Rhyme and rhythm together can be restrictive or prescriptive forces acting on a medieval poet. They can force poets to write a verse slightly different than the original thought in order to fill or fit the meter, or complement the rhyme. Thus the selected words were not only selected for sematic suitability, but for formal effect. They may not communicate the full truth (Mertens, 2005, p. 194) (Cramer, 1998, p. 180).

The computer model has difficulties with foreign words, outlier line lengths, and uncommon prosodic and metrical patternings. Belows are the lines about which the model was least confident, even though it correctly scanned (7), (9), and (10).:

(6) cordis | specu | la |tor[55]   (preferred)

´× × | ´× ×| — | `× ^

cor| dis spe| cu la| tor   (model)

—| `× ×| ´× ×| `× ^

(7) dâ | hiez sî | ûf | gân[56]

— | ´× ×| —| ´× ^

(8) nû | râtet | mir | alle durch | got[57]   (preferred)

× | ´× ×| — | ´× ◡ ◡ | ´× ^

[52]"There was a knight so learned," l. 1. Average marginal probability: 0.9991.
[53]"I know well that he himself confirms," l. 1162. Average marginal probability: 0.9988.
[54]"He is a very much nonsensical fool," l. 725. Average marginal probability: 0.9987.
[55]"The one who seeks the heart," l. 1357. Average marginal probability: 0.247.
[56]"She was ordered to go on top (of the table)." l. 1206. Average marginal probability: 0.274.
[57]"Now give me advice you all through God," l. 1482. Average marginal probability: 0.496.

TABLE 6 | Example (6), average probability 0.247.

Syllable	Value	Marginal prob.
cor	–	0.124
dis	`	0.132
spe	×	0.134
cu	´	0.204
la	×	.342
tor	`	0.549

TABLE 7 | Example (7), average probability 0.274.

Syllable	Value	Marginal prob.
dâ	—	0.028
hiez	´	0.114
sî	×	0.114
ûf	—	0.114
gân	´	0.998

nû | râtet | mir al | le durch | got   (model)

× | ´× ×| ´× ×|´× × | ´× ^

(9) ich en| kun| de ze | saler | ne[58]

× × | —| `× ×| ´× ×| `× ^

(10) dem ist | ouch | niht ze | wol[59]

´× ×| — | ´× ×| ´× ^

(11) einen | fremeden | tôt | niht ver|tragen[60] (preferred)

× ×| ´× ◡ ◡ | — | ´× × | ◡ ◡ ^

einen | freme| den tôt | niht ver| tragen   (model)

× × | ´× ×| ´× × | ´× × | ◡ ◡ ^

While the Latin in example 6 does not pose any problem for a human scanner trained as a medievalist, the model cannot identify the long syllable in specu*lā*tor, and having learned MHG, it would never presume a stressed syllable on the third syllable of a multi-syllabic word[61]. The marginal probabilities in **Table 6** are striking compared even to the other difficult lines, as the model is unsure about nearly every syllable[62]. The other troubling cases are more relatable. In example 7, we are confronted with the minimum number of syllables that a MHG epic verse is permitted to contain, one which the model nearly guesses correctly. Adding to the difficulty is that each word is monosyllabic, and as documented above, the model prefers bisyllabic words as double morae. The double morae are also

[58]"I could not to Salerne." l. 1018. Average marginal probability: 0.504.
[59]"He is not doing so well." l. 600. Average marginal probability: 0.505.
[60]"To bear another death?" l. 1329. Average marginal probability: 0.516.
[61]However, were the long ā marked it may have scanned the line correctly.
[62]In fact, using these probabilities, it is likely a promising task to be able to determine whether a word is MHG or not.

TABLE 8 | Example (8), average probability 0.496.

Syllable	Value	Marginal prob.
nû	×	0.921
râ	´	0.644
tet	×	0.381
mir	´	0.151
al	×	0.137
le	´	0.310
durch	×	0.427
got	´	0.999

TABLE 9 | Example (9), average probability 0.504.

Syllable	Value	Marginal prob.
ich	×	0.768
en	×	0.631
kun	—	0.379
de	`	0.348
ze	×	0.263
sa	´	0.473
ler	×	0.505
ne	´	0.668

TABLE 10 | Example (10), average probability 0.505.

Syllable	Value	Marginal prob.
dem	´	0.014
ist	×	0.014
ouch	—	0.013
niht	´	0.997
ze	×	0.997
wol	´	0.997

TABLE 11 | Example (11), average probability 0.516.

Syllable	Value	Marginal prob.
ei	×	0.791
nen	×	0.956
fre	´	0.309
me	×	0.166
den	`	0.167
tôt	×	0.233
niht	´	0.411
ver	×	0.420
tra	˘	0.712
gen	˘	0.996

TABLE 12 | Model's ease of scanning *Vierheber* texts based on line marginal probabilities.

Text	Median	Mean
Reinfried von Braunschweig	0.984907	0.952727
Der Schwanritter	0.984840	0.954346
Das Turnier von Nantes	0.984484	0.953066
Alexius	0.983046	0.953657
Herzmaere	0.982992	0.956888
Heinrich von Kempten	0.982073	0.948779
Pantaleon	0.981337	0.944830
Der Trojanische Krieg	0.981231	0.945008
Engelhard	0.980778	0.947011
Silvester	0.980702	0.948039
Der guote Gêrhart	0.978524	0.939438
Frauendienst (Bechst.) (Epik, Bechstein)	0.977328	0.934517
Herzog Ernst (Hs.D, strophig)	0.976312	0.926170
Alexander (R. v. E.) (Rudolf von Ems)	0.974599	0.930398
Barlaam und Josaphat	0.969378	0.926035
Frauendienst (Büech.) (Büechlîn, Bechstein)	0.966405	0.923810
Meleranz	0.965263	0.915054
Wigalois, der Ritter mit dem Rade	0.963911	0.919376
Gauriel von Muntabel	0.962624	0.911146
Biterolf und Dietleib	0.962019	0.917653
Helmbrecht	0.961616	0.917587
Tristan (H.v.F.)	0.960467	0.913434
Tandareis und Flordibel	0.958841	0.912529
Der arme Heinrich	0.956336	0.911926
Parzival	0.950174	0.911735
Dietrich und Wenezlan	0.948923	0.907889
Gregorius	0.948497	0.904120
Iwein	0.947784	0.902092
Lanzelet	0.947157	0.899856
Herzog Ernst (Hs. B)	0.946568	0.903171
Alexander (U.v.E) (Ulrich von Eschenbach)	0.945071	0.903506
Tristan (Ulrich v. Türheim)	0.944975	0.899818
Walberan	0.943629	0.897919
Laurin	0.943434	0.900584
Dietrichs Flucht	0.942704	0.898055
Willehalm (Wolfram)	0.942339	0.904198
Flore und Blanscheflur	0.942235	0.899201
Eneide	0.936819	0.889650
Karl der Grosse	0.935828	0.898457
Der Welsche Gast	0.934193	0.898327
Der Schlegel	0.933589	0.899505
Alexander Anhang	0.932088	0.887358
Erec	0.931306	0.890927
Daniel von dem blühenden Tal	0.922973	0.884164
Willehalm (U.v.T.)	0.918577	0.893880
Lambrechts Alexander (Strassburger Hs.)	0.894552	0.864145

only two characters (or phonemes) in length, an adverb and a preposition. **Table 7** shows that the model is not confident about any of the assignments except for the very last syllable, which as a monosyllabic ultima, is likely a stressed mora.

Example 8 with it's marginal probabilities in **Table 8** appears as classic trochaic tetrameter following the form of an easy prediction for the model, yet the natural stress of the bisyllabic "alle" prevents this scansion, forcing "mir" to be in a stressed

TABLE 13 | Ratio of double morae syllables.

Text	% double morae
Lambrechts Alexander (Strassburger Hs.)	0.097617
Eneide	0.095105
Flore und Blanscheflur	0.088926
Erec	0.080958
Daniel von dem blühenden Tal	0.079729
Herzog Ernst (Hs. B)	0.078354
Gregorius	0.077372
Lanzelet	0.077010
Helmbrecht	0.076015
Karl der Grosse	0.074808
Der arme Heinrich	0.074564
Alexander Anhang	0.072936
Dietrichs Flucht	0.071523
Iwein	0.070903
Reinfried von Braunschweig	0.069159
Gauriel von Muntabel	0.068819
Dietrich und Wenezlan	0.063602
Das Turnier von Nantes	0.063528
Der Schlegel	0.062627
Biterolf und Dietleib	0.060387
Tristan (Ulrich v. Türheim)	0.060161
Laurin	0.058949
Alexander (U.v.E) (Ulrich von Eschenbach)	0.057748
Walberan	0.056834
Willehalm (Wolfram)	0.052707
Alexander (R. v. E.) (Rudolf von Ems)	0.051154
Wigalois, der Ritter mit dem Rade	0.050662
Der Schwanritter	0.049955
Meleranz	0.049557
Der Trojanische Krieg	0.048831
Engelhard	0.048730
Pantaleon	0.047476
Parzival	0.046408
Tandareis und Flordibel	0.045926
Tristan (H.v.F.)	0.045665
Silvester	0.045198
Barlaam und Josaphat	0.043626
Herzmaere	0.042582
Herzog Ernst (Hs.D, strophig)	0.041741
Willehalm (U.v.T.)	0.040853
Frauendienst (Büech.) (Büechlîn, Bechstein)	0.040585
Alexius	0.040388
Der guote Gêrhart	0.040166
Heinrich von Kempten	0.038583
Der Welsche Gast	0.037997
Frauendienst (Bechst.) (Epik, Bechstein)	0.006331

position in the preferred scansion. The model is least confident about the three syllables it in fact scans incorrectly, ("mir al-le") but evidently believes the alternation to be the stronger choice than a midline, monosyllabic double mora followed by a divided lift, which itself is a rare occurrence.

TABLE 14 | Double morae word frequency.

Text	Double morae
Alexander (R. v. E.) (Rudolf von Ems)	rîche, persî, maere, waere, wîgant
Alexander (U.v.E) (Ulrich von Eschenbach)	rîche, waere, maere, swaere, daz
Alexander Anhang	alsô, daz, mit, stat, rîche
Alexius	eufêmiân, ougen, reine, sêre, lougen
Barlaam und Josaphat	lêre, rîche, sêre, mêre, arbeit
Biterolf und Dietleib	maere, waere, wîgant, rîche, rîchen
Daniel von dem blühenden Tal	waere, niht, daz, sô, alle
Das Turnier von Nantes	engellanden, reine, handen, guldîn, gesteine
Der Schlegel	alten, behalten, wære, besunder, ende
Der Schwanritter	brâbant, hiute, swaere, beiden, maere
Der Trojanische Krieg	waere, rîche, stunden, werden, reine
Der Welsche Gast	mêre, solde, wârheit, niht, sinne
Der arme Heinrich	heinrich, herre, wære, arbeit, güete
Der guote Gêrhart	rîche, güete, quote, waere, muote
Dietrich und Wenezlan	bôlân, strîten, rîten, berne, zîten
Dietrichs Flucht	maere, dietrîch, bernaere, berne, gerne
Eneide	rîche, wâre, troiân, mâre, turnûs
Engelhard	werden, maere, erden, waere, engelhart
Erec	waere, êrec, herre, mêre, êre
Flore und Blanscheflur	waere, sô, daz, mêre, swaere
Frauendienst (Bechst.) (Epik, Bechstein)	helm, heinrîch, mîn, alsô, sîn
Frauendienst (Büech.) (Büechlîn, Bechstein)	êre, mêre, sêre, swaere, maere
Gauriel von Muntabel	waere, maere, walbân, rîche, gâwân
Gregorius	waere, swaere, quote, herre, maere
Heinrich von Kempten	heinrich, truhsaeze, kempten, hende, zîten
Helmbrecht	helmbreht, gotelinde, maere, waere, muoter
Herzmaere	cleine, frouwen, sinne, beide, schouwen
Herzog Ernst (Hs. B)	rîche, lande, wîgant, maere, waere
Herzog Ernst (Hs.D, strophig)	daz, strîten, maere, liute, swaere
Iwein	waere, êre, arbeit, îwein, mêre
Karl der Grosse	ruolant, heiden, êre, sêre, rîche
Lambrechts Alexander (Strassburger Hs.)	alexander, rîche, wâren, dô, daz
Lanzelet	waere, maere, êre, wîgant, artûs
Laurin	laurîn, dietrîch, wîgant, dietleip, berne
Meleranz	rîche, zîten, wîgant, waere, maere
Pantaleon	werden, erden, ougen, heiden, worden
Parzival	gâwân, rîche, maere, waere, minne
Reinfried von Braunschweig	sinne, minne, wære, rîche, mære
Silvester	werden, erden, reine, waere, haete
Tandareis und Flordibel	rîche, wîgant, maere, waere, êre
Tristan (H.v.F.)	minne, herzen, marke, sinne, küneginne
Tristan (Ulrich v. Türheim)	maere, waere, ýsôt, minne, sôte
Walberan	laurîn, gerne, waere, berne, dietrîch
Wigalois, der Ritter mit dem Rade	manheit, waere, korntîn, rîche, maere
Willehalm (U.v.T.)	niht, markîs, waere, daz, minne
Willehalm (Wolfram)	rîche, waere, markîs, minne, heimrîch

While the model correctly predicts example 9 (see marginal probabilities in **Table 9**), it has little confidence in the middle of the line. A two syllable anacrusis is not uncommon, though not frequent, and the two light, open syllables "de" as a stressed syllable and "ne" as a final stressed syllable further add to confusion. Although to give the model the benefit of the doubt, "salerne" is a proper noun.

The model also correctly predicts example 10 (see it's marginal probabilities in **Table 10**), though again MHG prosody would not suggest "dem" or "ouch" as double morae. The duration of the line is typical and the marginal probabilities confirm this. It is particularly striking in this example that the model is very uncertain about the beginning of the line. While "dem" is clearly a rare double mora, in this position it would even be considered a rare stress, and would typically be scanned as part of the anacrusis. In contrast to the beginning of the line, the model has great confidence in the end of the line, where "ze" would commonly be unstressed (often even elided or in a divided lift), and the heavy syllables "niht" and "wol" are assigned the remaining stresses.

Example 11 shows another outlier in terms of line length (10 syllables, see it's marginal probabilities in **Table 11**). Whether or not we accept "fremeden" as three syllables or two ("fremden" is attested in manuscript A) the line begins with a two syllable anacrusis (notably, only one word), and ends in a three syllable word, the first syllable of which is unstressed. Once again, the syllables about which the model is least certain are also those incorrectly scanned. Yet the model evidently believes a double mora for "tôt" and a divided lift are less likely than retaining alternation. The masculine bisyllabic cadence, particularly difficult to scan on first read, is correctly identified.

These marginal probabilities additionally allow us to calculate the difficulty of scanning any given MHG *Vierheber* text as a whole (and even specific sections of any text). Although we are determining difficulty for the model, we have shown above that the model appears to approximate a human annotator. Thus, difficulty for the model may be reasonably inferred to represent difficulty for a human annotator. This can be computed by taking the average of each syllable's maximum marginal probability over the syllables in a line. In this sense, each syllable has a marginal probability for each possible metrical value[63]. The examples illustrate that the typical trochaic tetrameter causes problems for neither the model nor the human scanner, while unexpected double morae and longer anacrusis are cause to stop and think, particularly when syllables with relatively few phonemes, or monosyllabic words, are properly assigned double mora value. To sort the epic *Vierheber* texts in the Mittelhochdeutsche Begriffsdatenbank (MHDBDB)[64] corpus by difficulty of the meter, the median and mean marginal probabilities for all lines in each text are calculated (**Table 12**). While Konrad von Würzburg and the anonymous author of Reinfried von Braunschweig utilize

---

[63]$marginal(t, s)$ is the marginal probability of tag t for syllable s, and thus $\sum_{t'} marginal(t', s) = 1$. For a line with N syllables, the average of each syllable's maximum marginal probability over the syllables in a line is thus $\frac{1}{N} \sum_{i=1}^{N} \arg\max_t marginal(t, w_i)$.

[64]Springeth et al. (1992-2017).

TABLE 15 | Shannon entropy for 1,000 random line samples.

Text	Foot	Line
Eneide	1.252917	3.763736
Lambrechts Alexander (Strassburger Hs.)	1.194705	3.980579
Iwein	1.167059	3.405163
Daniel von dem blühenden Tal	1.140783	3.498383
Helmbrecht	1.099641	3.039134
Der arme Heinrich	1.094388	3.217970
Laurin	1.030382	3.407795
Erec	1.005811	3.294478
Gregorius	0.980259	3.133976
Willehalm (Wolfram)	0.979950	3.159355
Tristan (Ulrich v. Türheim)	0.971779	3.185109
Alexander Anhang	0.969258	3.248891
Karl der Grosse	0.968051	2.983325
Der Schlegel	0.934412	3.277041
Meleranz	0.922810	2.891814
Barlaam und Josaphat	0.917566	2.968009
Flore und Blanscheflur	0.917468	3.115846
Herzog Ernst (Hs. B)	0.914392	3.191947
Gauriel von Muntabel	0.906261	3.041430
Tristan (H.v.F.)	0.903151	3.103300
Dietrichs Flucht	0.894119	3.041568
Walberan	0.884103	3.246322
Willehalm (U.v.T.)	0.881089	3.078347
Heinrich von Kempten	0.875890	2.665526
Parzival	0.868807	3.217873
Alexander (U.v.E) (Ulrich von Eschenbach)	0.867947	2.963053
Tandareis und Flordibel	0.867046	2.642869
Biterolf und Dietleib	0.854046	2.824735
Dietrich und Wenezlan	0.852161	2.974738
Pantaleon	0.850183	2.725092
Lanzelet	0.848197	3.022754
Alexius	0.845937	2.837503
Das Turnier von Nantes	0.809099	2.304515
Der Trojanische Krieg	0.803099	1.792531
Engelhard	0.800056	1.944227
Herzmaere	0.770970	2.488530
Alexander (R. v. E.) (Rudolf von Ems)	0.768860	2.413441
Der Schwanritter	0.761848	2.485584
Der Welsche Gast	0.710632	2.645962
Silvester	0.703183	2.412775
Frauendienst (Büech.) (Büechlîn, Bechstein)	0.690972	2.655543
Herzog Ernst (Hs.D, strophig)	0.666231	2.460030
Der guote Gêrhart	0.624613	2.388511
Wigalois, der Ritter mit dem Rade	0.622915	2.714158
Reinfried von Braunschweig	0.621006	1.908564
Frauendienst (Bechst.) (Epik, Bechstein)	0.303572	1.881531

the ringing cadence frequently, Ulrich von Liechtenstein does not, yet both do so consistently and in a predictable manner that distinguishes them from the rest of the texts. The major use of simple trochaic tetrameter, or double morae only in the penultimate foot, likely further pushed texts toward the top of

the list (evidenced by the model's ease in predicting strictly trochaic verse). An odd mix of texts appear at the bottom of **Table 12**. Both Ulrich von Türlin's *Willehalm*, or *Arabel*, and Wolfram's *Willehalm* evidently employ more difficult metrical schemata, as does Der Stricker's *Daniel von dem blühenden Tal*, perhaps a consequence of his unique *Strickerkadenz*. Particularly interesting is the difficulty of scanning *Der Welsche Gast*. Because Thomasîn was not writing in his native language, his native language being Italian, perhaps he more often confuses stress in MHG and thus creates difficulties for the model despite a very simple trochaic patterning. While this ranking is interesting for scholars and those teaching MHG meter, it cannot be fully validated as such a measure has not yet been considered in the scholarship. Nevertheless, assuming preference for trochaic tetrameter in easier meters this ranking is roughly accurate.

10. DOUBLE MORA

Both the analysis above and the following analyses assume a high degree of reliability in the model to accurately scan poetry from different authors than those in the train and test sets. This certainly harbors potential bias. The training and testing data authors were chosen for their generality and the degree to which it is believed they were imitated by other authors in the corpus. While additional data from each author were not annotated, these authors all followed the same principles and framework as outlined in Heusler's theory. The features given to the model were decisive for all authors of MHG epic poetry. However, it is certainly possible other authors viewed word stress differently due to geographic or other differences. Future work should validate the model's performance on a variety of MHG epic poetry texts.

Perhaps the most interesting aspect of MHG verse is the *beschwerte Hebung*, or double mora. Double morae allow for a syllable to carry twice the metrical weight (and likely thus twice the duration) of a normal syllable. The use of double morae is often considered to be an important stylistic choice in indicating semantically, or contextually, important words in a verse: "only then did the monosyllabic foot become a "declamatory machinery"; it exhilaratingly disrupted the up and down pacing and increasingly yielded the natural stress of the language" (Heusler, 1956, p. 118). Since a double mora occupies an entire foot, the following syllable is necessarily stressed, since it is the beginning of the next foot. Hence the disruption of the otherwise naturally occuring alternation. Heusler also notes that double morae were quite common in early MHG epics such as Heinrich von Veldeke's *Eneide*, and even had an "altertümlich" ("antiquated"), effect in later works containing many double morae, such as Hartmann's *Erec* and *Gregorius* (Heusler, 1956, p. 101).

The trend began to move away from this "Germanic" characteristic toward adopting a more strict trochaic style from the French tradition, seen in Rudolf von Ems' *Alexander* and nearly eliminated by Ulrich von Liechtenstein and Konrad von Würzburg, who aim for almost exclusively strict stress alternation (Heusler, 1956, pp. 101, 118). Heusler understands Konrad as the

endpoint, after whom German verse began to be written more freely once again. **Table 13** confirms most of these observations. Of these texts, the five most common double morae are: "rîche" ("rich," "powerful"), "waere" ("would be"), "maere" ("story"), "ere" ("honor"), "mêre" ("more") (closely followed by "sêre" ("very")). All double morae must be heavy phonologically, but it is interesting that all of the most frequently occurring double morae have *front* vowels, and all but rîche have an open syllable (the velar fricative "ch" is construed as ambisyllabic in MHG). Another phonological similarity among them is that each begins with a *sonorant* consonant (/m/, /n/, /w/, /j/, /l/, /r/)[65], and all, except for "rîche," end in "re." Sonorant consonants are voiced consonants with continuous airflow faciliated by minimal obstruction in the vocal tract. This is in contrast with obstruent consonants, such as stops (e.g., /p,b/) or fricatives (e.g., /f,v/), in which a greater occlusion in the oral cavity occurs. Furthermore, sonorants are typically louder than obstruents. In preferring sonorant consonants in double mora syllables, MHG poets clearly aimed for the greatest resounding word to place in the double morae position, perhaps pointing to a greater significance of the vocal performance tradition in MHG epic verse than has previously been acknowledged. Since open syllables allow for an unobstructed flow of air from the mouth, they may be preferred for notes of extended length (or melismatic syllables), especially considering the relative dearth of open, heavy syllable words in MHG. **Table 14** shows the most common double mora for each text. It is evident that double morae contribute heavily to the characterization of a given text, which can come as no surprise given that double morae would present themselves most prominently in memory. These are also concepts that are often studied for their respective text, such as "ere," "maere," and "rîche."

11. DIVERSITY IN *VERSFÜLLUNG*

Heusler was particularly interested in the variety of ways a foot could be filled with metrical values. To measure Heusler's diversity of "Versfüllung" (filling of the verse) in its entirety, we adopt a popular measure of diversity originally developed in the context of information theory and widely used in environmental science, namely Shannon entropy (Shannon, 1948-07). Shannon entropy fits this application well, considering that Shannon's original problem dealt with string prediction given a set of characters. The entropy measure quantifies the degree of certainty in predicting a random character from a string dataset, given a finite number of characters. Similarly, a useful measure of "Versfüllung" entails the degree of certainty in predicting a specific foot or line. To this end, we calculate the Shannon entropy for a complete foot (excluding anacrusis and the last foot) and for the entire line, sampling 1,000 feet with replacement from each text. The results are shown in **Table 15**. The earlier, more "Germanic" texts display a wider diversity in metrical patterning, while Ulrich von Liechtenstein and Konrad

---

[65]Although the pronunciation of MHG "w" is not certain; it may have been a labiovelar glide, phonetically [w], a sonorant, or perhaps the labiodental fricative [v] as it is in NHG, an obstruent consonant.

von Würzburg aim toward monotonous trochees. We also see all the works attributed to Hartmann von Aue in the top ten for diversity. Save *Parzival* and *Der Welsche Gast*, the top ten contains the works most often discussed by scholars.

12. CONCLUSION

This paper has presented a new approach to a very old problem for medieval German scholarship. Investigating the meter of this tradition poses unique challenges to literary scholars, philologists, and computational linguists alike. By constructing a supervised model of the meter, this paper demonstrates the benefits of a quantitative corpus-wide analysis enabling us to characterize its idiosyncrasies and suggest improvements to the current pedagogical approach. Moreover, this approach has proven generalizable to other traditions, having been taken up by scholars working on other languages with the results presented here as baseline[66]. A promising extension of this work would be a comparative analysis, particularly in the medieval context, as

---

[66]See Agirrezabal et al. (2016), Navarro (2017), and Mittmann (2016).

it is well-known that these traditions were influenced by their contemporaries.

REFERENCES

Agirrezabal, M., Alegria, I., and Hulden, M. (2016). "Machine learning for metrical analysis of english poetry," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (Osaka), 772–781.

Agirrezabal, M., Arrieta, B., Astigarraga, A., and Hulden, M. (2013). "ZeuScansion: a tool for scansion of English poetry," in *Finite State Methods and Natural Language Processing* (St Andrews), 18.

Bögl, H. (2006). *Abriss der Mittelhochdeutschen Metrik: Mit Einem übungsteil.* Hildesheim: Georg Olms Verlag.

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing With Python: Analyzing Text With the Natural Language Toolkit.* O'Reilly Media, Inc.

Braune, W., and Ebbinghaus, E. A. (1994). *Althochdeutsches Lesebuch.* Tübingen: Niemeyer.

Brill, E. (1995). Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging. *Comput. Linguist.* 21, 543–565.

Cramer, T. (1998). *Waz Hilfet Âne Sinne Kunst?: Lyrik im 13. Jahrhundert: Studien zu Ihrer Ästhetik. Vol. 148.* Erich Schmidt Verlag GmbH & Co KG.

Dimpel, F. M. (2004a). *Computergestützte Textstatistische Untersuchungen an Mittelhochdeutschen Texten-Bd. II.* Daten und Programme.

Dimpel, F. (2004b). Textstatistische analysen an mittelhochdeutschen texten. *Jahrbuch Computerphilol.* 6, 95–118.

Dimpel, F. (2015). Automatische mittelhochdeutsche metrik 2.0. *Phil. Netz* 73, 1–26.

Domanowski, A., Rauert, Y., Rüther, H., and Tomasek, T. (2009). *Mittelhochdeutsche Metrik Online.* Available online at: https://www.uni-muenster.de/MhdMetrikOnline/ (Accessed 04 02, 2016).

Estes, A., and Hench, C. (2016). "Supervised machine learning for hybrid meter," in *Proceedings of the Fifth Workshop on Computational Linguistics for Literature* (San Diego, CA), 1–8.

Fox, A. (2000). *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals.* New York, NY: Oxford University Press.

Giegerich, H. J. (1985). *Metrical Phonology and Phonological Structure: German and English.* Cambridge: Cambridge University Press, 1-301

Greene, E., Bodrumlu, T., and Knight, K., (2010). "Automatic analysis of rhythmic poetry with applications to generation and translation," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language*

Processing. Association for Computational Linguistics (Cambridge, MA), 524–533.

Hartman, C. O. (1996). *Virtual Muse: Experiments in Computer Poetry.* Hanover, NH: Wesleyan University Press.

Hartmann, v. A., and Mertens, v. V. (2005). *Der arme Heinrich. Bibliothek des Mittelalters.* Cambridge, UK: Ann Arbor, MI: Chadwyck-Healey; ProQuest Information and Learning. Available online at: http://uclibs.org/PID/108047

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguist. Inq.* 20, 253–306.

Hench, C. (2017). "Phonological soundscapes in medieval poetry," in *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature,* (Berkeley, CA), 46–56.

Heusler, A. (1956). *Deutsche Versgeschichte: Mit Einschluss Des Altenglischen Und Altnordischen Stabreimverses. Grundriss Der Germanischen Philologie 8.* Berlin: W. De Gruyter.

Karg-Gasterstädt, E. (1925). *Zur Entstehungsgeschichte des Parzival Vol. 2.* M. Niemeyer.

Kohrs, K. H. (1969). Zum verhältnis von sprache und musik in den liedern neidharts von reuental. *Deutsche Vierteljahrsschrift Literaturwissenschaft Geistesgeschichte* 43:604.

Kuhn, H. (1969). *Text Und Theorie. His Kleine Schriften, Bd. 2.* Metzler: Stuttgart.

Lafferty, J., McCallum, A., and Pereira, F. C. N. (2001). "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in *ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning,* (Morgan Kaufmann Publishers Inc. San Francisco, CA).

Longfellow, H. W. (1932). *Poems, Including Evangeline, The Song of Hiawatha, The Courtship of Miles Standish, Tales of a Wayside Inn.* No. 56. Modern library.

Mathisen, R. W. (2003). *People, Personal Expression, and Social Relations in Late Antiquity.* Ann Arbor, MI: University of Michigan Press.

McAleese, G. (2007). *Improving Scansion With Syntax: An Investigation Into the Effectiveness of a Syntactic Analysis of Poetry by Computer Using Phonological Scansion Theory.* PhD diss., Ph. D. thesis, Open University.

Mertens, V. (2005). "Was ist Rhythmus in Minnesang?" in *Aus dem Takt: Rhythmus in Kunst, Kultur und Natur,* (Bielefeld: Transcript), 175–198.

Mittmann, A. (2016). *Escansão Automática de Versos em Português.* Ph.D. thesis, Universidade Federal de Santa Catarina.

März, C. (1999). "Metrik, eine wissenschaft zwischen zählen und schwärmen?," in *Mittelalter: Neue Wege Durch Einen Alten Kontinent,* eds J.-D. Müller and H. Wenzel (Hirzel: S. Hirzel Verlag), 317–332.

Navarro, B. (2015). "A computational linguistic approach to Spanish Golden Age Sonnets: metrical and semantic aspects," in *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, (Alicante), 105–113.

Navarro-Colorado, B. (2017). A metrical scansion system for fixed-metre Spanish poetry. *Digit. Scholar. Human.* 33, 112–127.

Norberg, D. (2004). *An Introduction to the Study of Medieval Latin Versification*. CUA Press.

Okazaki, N. (2007). *CRFsuite: A Fast Implementation of Conditional Random Fields (CRFs)*. Available online at: http://www.chokkan.org/software/crfsuite/

Plamondon, M. R. (2006). Virtual verse analysis: analysing patterns in poetry. *Literary Linguist. Comput.* 21, 127–141. doi: 10.1093/llc/fql011

Shannon, C. E. (1948-07). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Springeth, M., Morocutti, N., and Schlager, D. (1992-2017). *Mittelhochdeutsche Begriffsyearnbank (mhdbdb)*. Universität salzburg. Available online at: http://www.mhdbdb.sbg.ac.at/ (Accessed 10, 01, 2016).

Tervooren, H. (1979). *Minimalmetrik: zur Arbeit mit Mittelhochdeutschen Texten Vol. 285*. Kümmerle.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.