

# The Parsed Corpus of Middle English Poetry

Richard Zimmermann



## 1. Definition and Motivation

The Parsed Corpus of Middle English Poetry (PCMEP) is a fully parsed and annotated corpus of Middle English verse texts.<sup>[1]</sup>

There is a substantial gap in English prose texts between c. 1250 and 1350. The main goal of the PCMEP is to help close this gap with available poetic records from the same period. In order to be able to assess the genre difference between prose and poetry, the corpus covers a slightly greater time span than that, namely c. 1150 to 1420 (Helsinki periods M1, M2, M3).

The PCMEP is parsed according to the same guidelines as its much larger sister corpus, the Penn-Parsed Corpus of Middle English, second edition (PPCME2).<sup>[2]</sup> Thus, researchers familiar with the PPCME2 do not have to learn any new annotation schemes and can use their PPCME2 search queries without modification on the PCMEP text files as well.

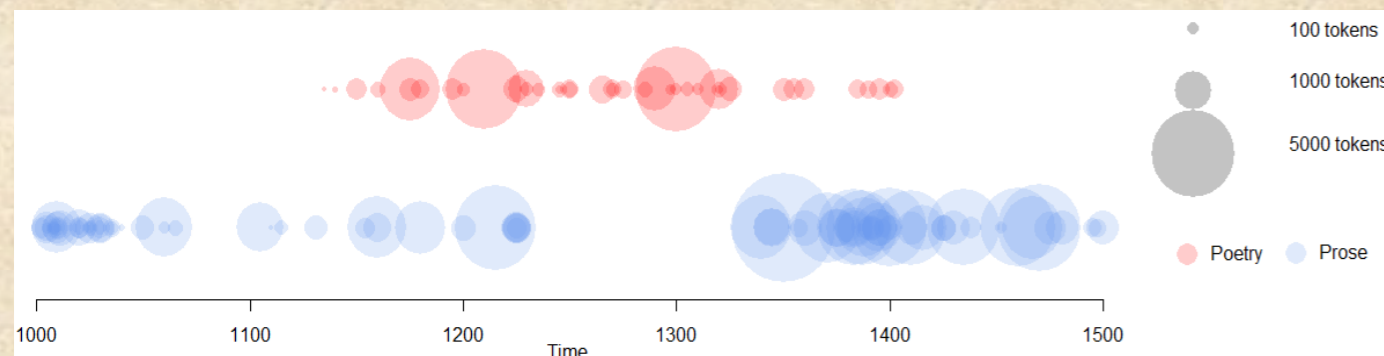


Figure 1: Temporal distribution of prose text and poetry corpora

## 2. Workflow

Appropriate texts are digitized from a large number of sources, followed by pre-processing, automatic parsing and finally manual correction.<sup>[3],[4]</sup>

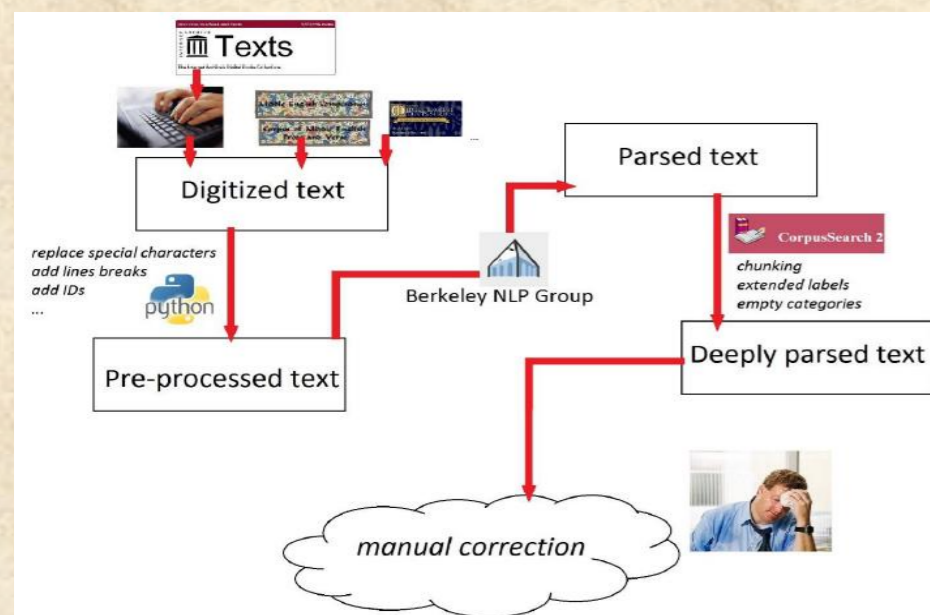


Figure 2: Illustration of the pipeline of tasks involved in the creation of parsed files

## 3. Example of a Parsed Tree

The current text (May 2018) to be parsed is *St. Margaret of Antioch* (MEC \*St.Marg.(2), IMEV 203, Manual 2.V.184(c)), composed c. 1220, 2'915 words. Horstmann (1881) serves as the base edition, out of copyright.

Olde ant yonge, i preie ou . oure folies for to lete;  
Denchet on god þat yef ou wit . oure sunnes to bete!

Figure 3: First line in edition

The parsed tree for the first sentence looks as follows:

```
(IP-MAT (LB | |))
(NP-VOC (ADJ olde) (CONJ ant) (ADJ yonge))
(, ,)
(NP-SBJ (PRO i))
(VBP preie)
(NP-OB2 (PRO ou))
(, ,)
(IP-INF (NP-OB1 (PRO$ oure) (NS folies))
(FOR for)
(TO to)
(VB lete))
(; .)
(LB | |))
(ID Margrete,489.1.1)
```

Figure 4: The parse tree for the first sentence in the St. Margaret file

The parsed trees can be retrieved with special query software according to a any number of conditions.<sup>[5]</sup>

```
opro-V.q      V-opro.q
define: cs.def  define: cs.def
node: IP*      node: IP*
query: (IP* idoms NP-OB*)
AND (NP-OB* idomonly PRO)
AND (IP* idoms
VBP|VBD|DOP|DOD|HVP|HVD)
AND (NP-OB* precedes
VBP|VBD|DOP|DOD|HVP|HVD)

query: (IP* idoms NP-OB*)
AND (NP-OB* idomonly PRO)
AND (IP* idoms
VBP|VBD|DOP|DOD|HVP|HVD)
AND
(VBP|VBD|DOP|DOD|HVP|HVD
precedes NP-OB*)
```

Figure 5: Example queries to find relative orders between object pronouns and finite verbs

## 4. Current Size and Goal

The corpus currently includes 42 Middle English poems with a total of 175'757 words. Detailed information on each text can be found on [www.pcme.net](http://www.pcme.net). The ultimate goal for the corpus is to reach a size of 1 million words.

Text Name	Estimated Date	Word Count
The First Worcester Fragment	1135	175
The Grave	1140	209
Body and Soul	1150	2'781
Pater Noster	1160	1'906
Poema Morale	1175	4'080
The Proverbs of Alfred	1180	3'095
Lord as Thou art one God	1195	2'910
A Good Orison of Our Lady	1205	1'501
A Lofsong to Our Lady (Wooing Group)	1210	722
A Lofsong of Our Lord (Wooing Group)	1210	1'754
An Orison of Our Lord (Wooing Group)	1215	1'274
Bastuary	1225	4'259
Wise Admonitions	1225	823
The Wooing of the Lord (Wooing Group)	1230	4'010
Maximian	1240	1'338
The Harrowing of Hell	1240	1'461
A Lutel Soth Sermon	1245	438
The Passion of Our Lord	1245	6'429
Love-Ron	1250	1'724
The Eleven Pains of Hell	1250	3'610
The Owl and the Nightingale	1250	10'941
Joseph and Jacob	1265	4'605
The Thrush and the Nightingale	1270	1'108
Bame Sirith	1270	2'539
The Fox and the Wolf	1275	1'816
A Metrical Treatise on Dreams	1285	1'736
Havok the Dane	1290	1'794
The Song of the Husbandman	1297	6'04
The Land of Cockayne	1300	1'081
Kyng Alisunder	1300	45'881
The Legend of Frideswide	1305	1'884
Adam Davy's Five Dreams	1310	1'155
Amis and Amiloun	1320	15'925
An Orison of the Five Joys	1320	1'560
The Life of Saint Marina	1320	1'384
The Simonie	1325	4'817
The Dispute between Mary and the Cross	1350	3'247
How to Hear Mass	1355	4'043
Wynnerre and Westoure	1360	4'743
Sir Cluges	1395	3'384
The Bird with Four Feathers	1400	1'508
The Letter of Cupid	1402	3'522
		Total: 175'757

Figure 6: Poetry texts included in the PCMEP as of May 2018

## 5. Example Application

Middle English replaces *then* with *when* as the subordinator in subordinate temporal clauses.<sup>[6],[7]</sup>

(1) a. ac [ba he was þæt ger biseop ]  
but then he was that year bishop  
he witgode þæt se hæland sceolde sweltan for ðære þeode  
he predicted that the Savior should die for that people  
"But when he was bishop that year, he prophesied that the Savior should die for the people"  
(cowsgosp.Jn\_[WSCp]:11.51.6768) (c. 990)  
b. but [whanne he was bischop of that geer, ]  
but when he was bishop of that year  
he prophesiede, that Jhesu was to die for the folc  
he prophesied that Jesus was to die for the folk  
'But when he was bishop that year, he prophesied that Jesus would die for the people'  
(CMNTEST.11.40J.1143) (c. 1383)

Example 1: Conservative *þa* is replaced with innovative *when*

The change can be measured with great precision in the PCMEP. Statistical analyses return a rate of change which can then be compared to prose texts.

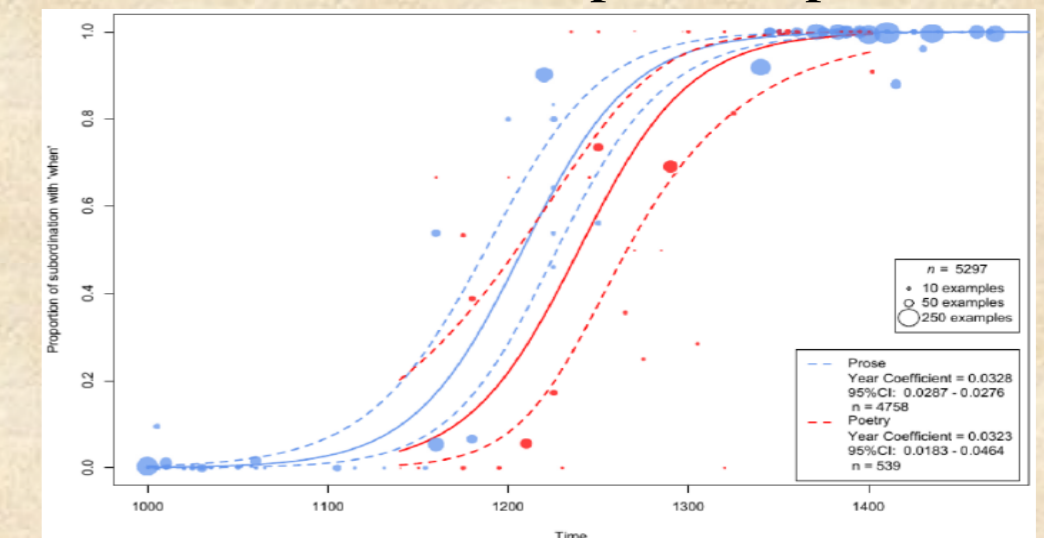


Figure 7: Development of temporal subordination with when

Poetry texts strengthen findings from prose texts.

## 6. Conclusion

The PCMEP closes a substantial gap in the availability of parsed, digitalized texts from the history of English. It allows fast, objective and accurate retrieval of linguistic data, and offers a new Middle English dataset to replicate previous findings and test new hypotheses.

### Acknowledgments

The PCMEP was supported by a Doc.Mobility grant from the Swiss National Science Foundation (P1GEP1\_148611).

## References

[1] Zimmermann, R. (2015) *The Parsed Corpus of Middle English Poetry*. University of Geneva. [2] Kroch, A. and Taylor, A. (2000) *Penn-Helsinki Parsed Corpus of Middle English*, University of Pennsylvania. [3] Petrov, S & Klein, D. (2007) *The Berkeley Parser* [4] Abney, S. (1991) 'Parsing by chunks' In: Berwick, R. and C. Tenny, eds, *Principle-Based Parsing*, pp. 257-278. [5] Randall, B. (2004) *CorpusSearch 2*. [6] Declerck, R (1997) *When-clauses and temporal structure*. London: Routledge. [7] Mitchell B. (1985) *Old English Syntax*. Oxford: Clarendon.