Reasoning About Genesis or: The Mechanical Philologist

Gregor Middell
Dietmar Pravida
Moritz Wissenbach

Goethe's "Faust"

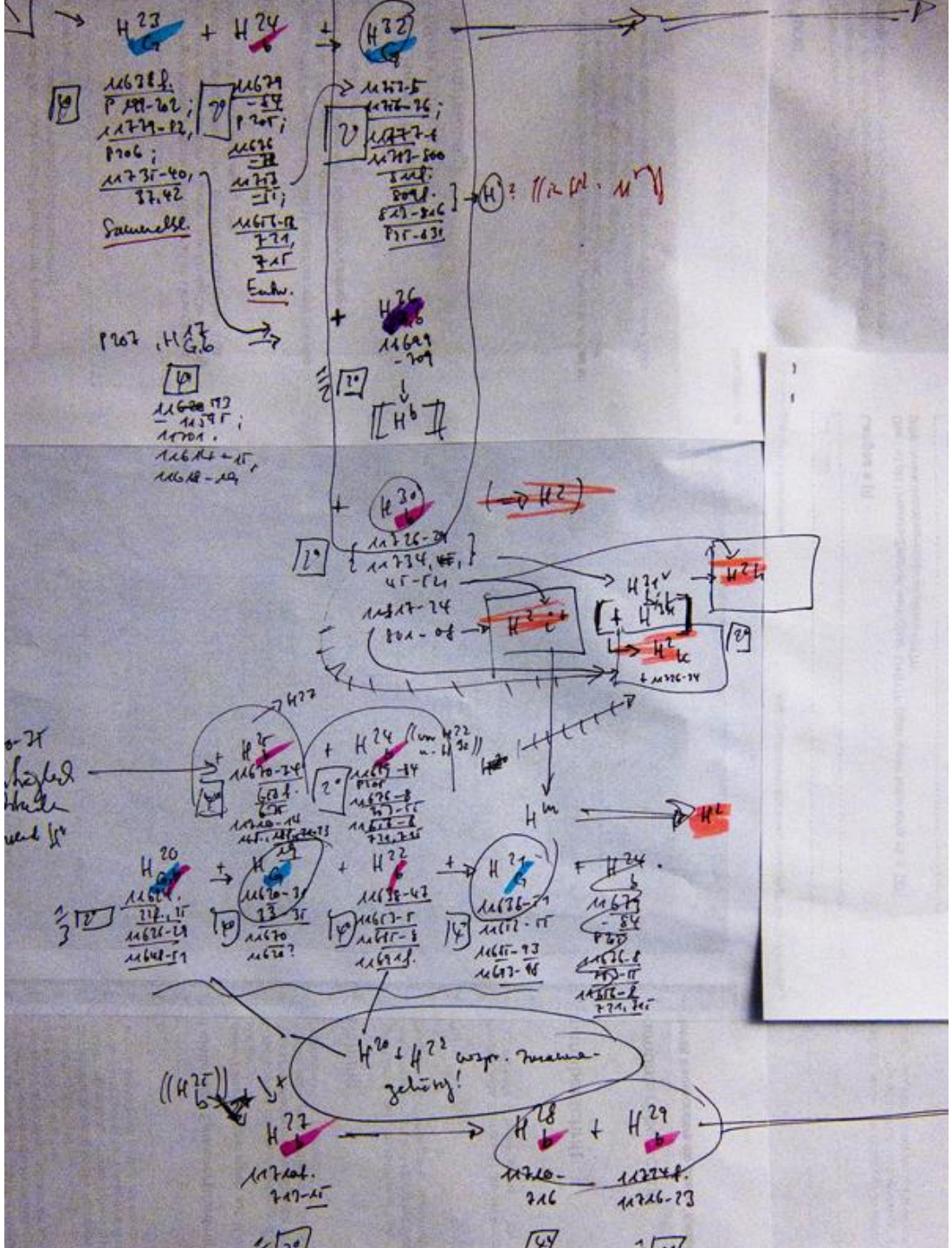
- ca. 1773–1831/32
- first (and to date only) critical edition 1887
- new critical edition, prepared by A. Bohnenkamp, F.Jannidis, S.Henke since 2009
- ca. 500 archival units/manuscripts extant

Reconstructing the genesis of "Faust"

- dating of individual manuscripts
- establishing relations between manuscripts
- taking account of a century of scholarship

Methods of dating

- methodological reflection
- relative – absolute chronology



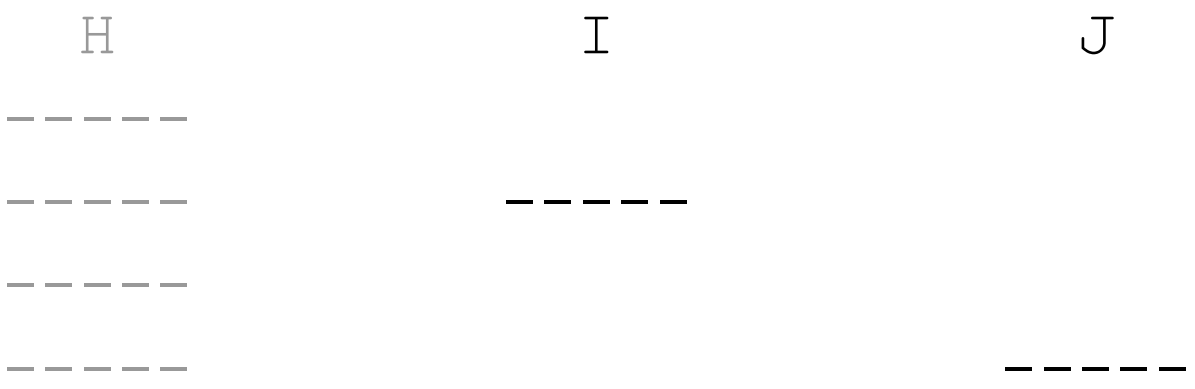
Evidence for dating manuscripts

- explicit dates on the ms.
- material properties (paper, ink, watermark etc.)
- external cues (mentions in diary, letter, table talk etc.)
- "logic" / dynamic of genesis inherent to the text

Computer-aided dating

- tool for editors
- provide the same tool and data to users
- task: formalizing the practice of dating using formal logic

Syntagmatic precedence (*syn*)

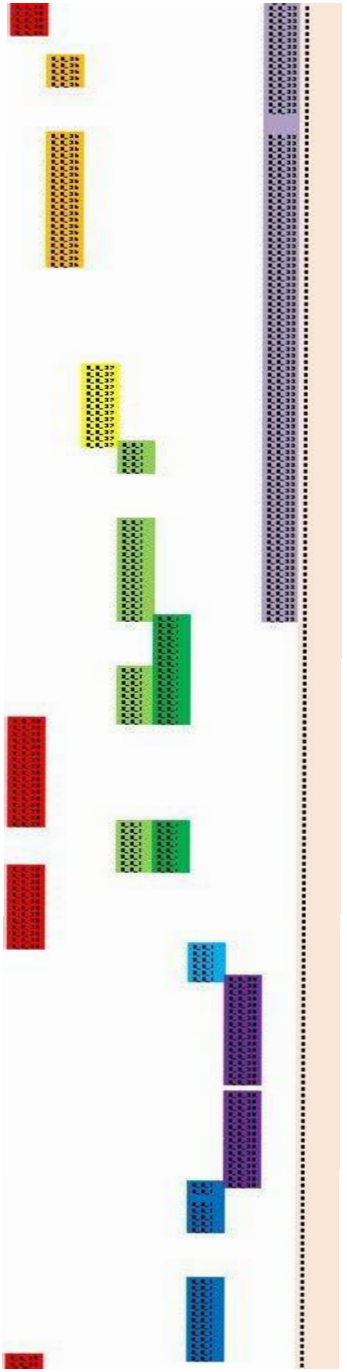


I *syntagmatically precedes* J or

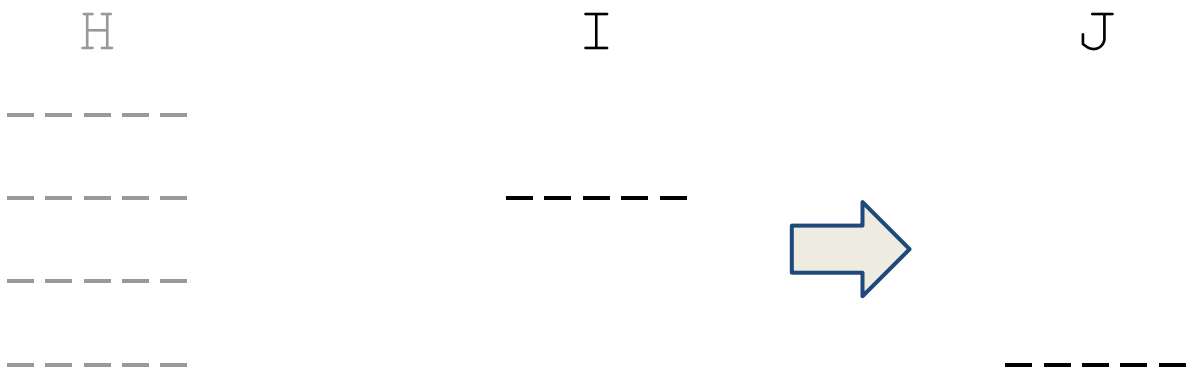
syn (I, J)

if J follows I with respect to a larger text H containing both

Last scene of "Faust" (Bergschluchten – Mountain Ravine)



Syntagmatic precedence (*syn*)



Rule *r_syn*:

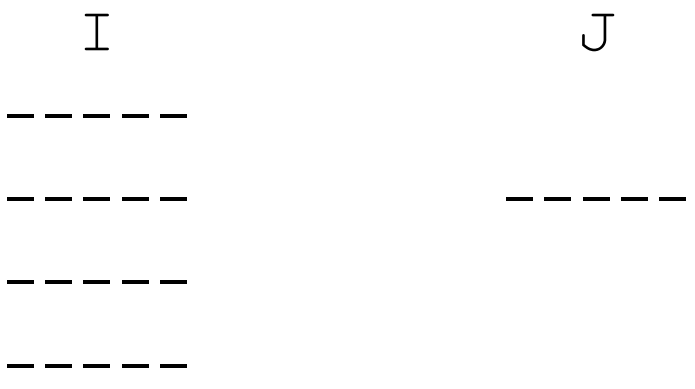
syn (I, J) ⇒ pre (I, J)

If I syntagmatically precedes J, then it is earlier than J

Syntagmatic precedence (*syn*)

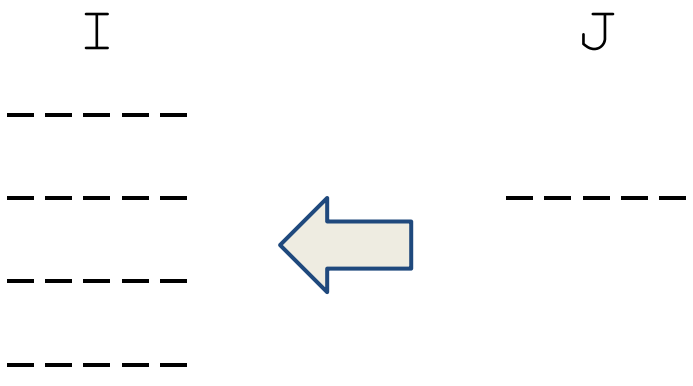
(graph)

Paradigmatic containment (*pcon*)



Paradigmatic containment: J is *paradigmatically contained* by I, if some part of J lies within the range of the overall syntagmatic interval of I and J is actually contained in I

Paradigmatic containment (*pcon*)



Rule *r_pcon*:

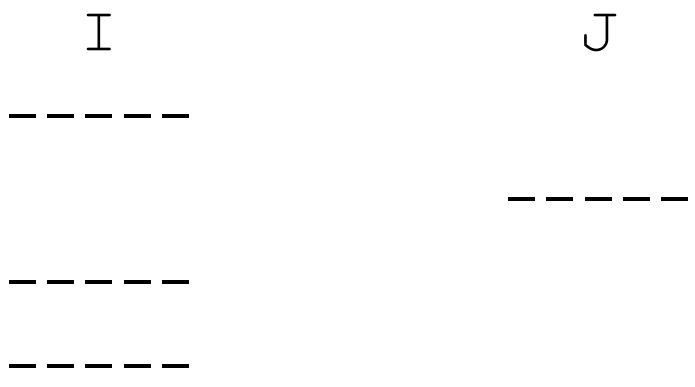
pcon (I, J) ⇒ pre (J, I)

If I paradigmatically contains J, then J is earlier than I

Paradigmatic containment (*pcon*)

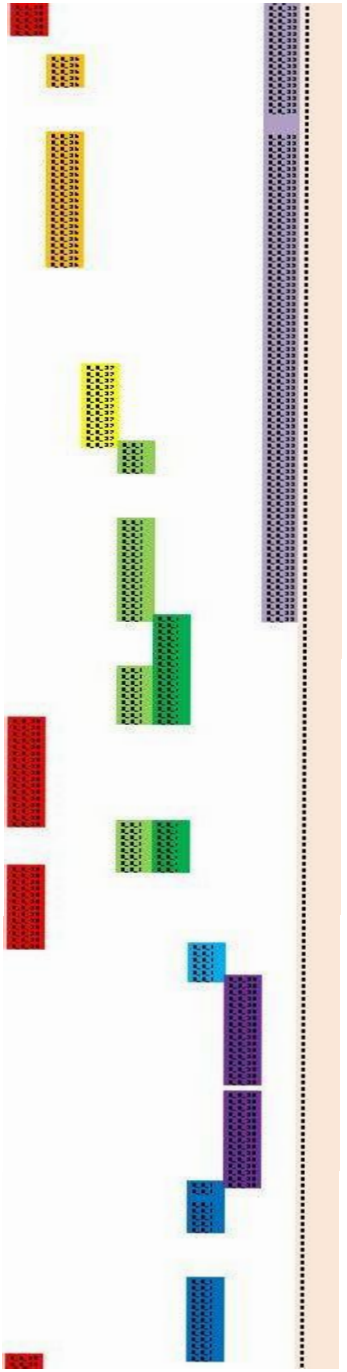
(graph)

Exclusive containment (*econ*)

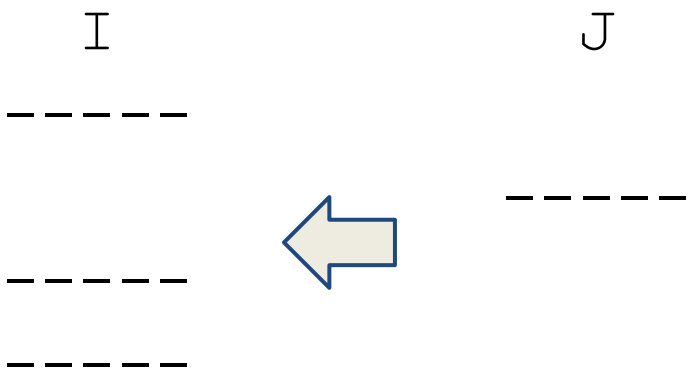


The overall syntagmatic interval J is *exclusively contained* by the overall syntagmatic interval of I, if some part of the text of J lies within the overall syntagmatic interval of I without being actually contained by I

Last scene of "Faust" (Bergschluchten – Mountain Ravine)



Exclusive containment (*econ*)



Rule *r_econ*:

econ (I, J) -> pre (I, J)

If I exclusively contains J, then it I is earlier than J

Knowledge from research

- collection of data from scholarly research on the dating of Faust manuscripts

(graph)

- new research
- overrule inferred relations

Application of Rules

- Transitivity

  $$\text{syn}(I, J) \wedge \text{syn}(J, K) \Rightarrow \text{syn}(I, K)$$

- Priorisation of Rules

  $$r\_syn < r\_pcon < r\_econ < r\_exp$$

- Combining

  (graph)

Evaluation of inference rules

- An inference rule is a formalized hypothesis about an authors writing habits
- Can be evaluted by comparison with established knowledge
- Measures
  - Coverage
  - Recall
  - Accuracy

Evaluation of inference rules

	Coverage [%]	Recall [%]	Accuracy [%]
r_syn	95.2	22.6	23.7
r_pcon	43.8	43.8	100.0
r_econ	3.4	2.7	80.0
r_syn_ante	95.2	72.6	76.3

Appendix

Outlook

- Interactive rule formulation & priorisation
- Application to act I-IV of "Faust"
- To other works / authors
- Induction of rules

Formalism, implementation

- First order predicate logic solver
  Too slow
- Description logic (OWL)
  Not expressive enough (no negation)
- Prolog
  No true negation

- Imperative implementation