

The Malmö University-Chalmers Corpus of Writing as a Process

Purposes

To build a **drafts corpus** to facilitate the study of writing processes from different perspectives; to study features, such as feedback, argumentation techniques and rhetorical structures; and to use the corpus for the teaching of academic writing

The corpus

Pilot version of 500,000 words made up of three drafts of:

- 400 student texts
- 50 PhD student texts
- Peer and teacher comments
- Self-reflective comments

Complete corpus: approximately 1.2 million words

Our contribution

The MUCH corpus:

- ✓ includes multiple drafts of student texts
- ✓ tags both writing and feedback processes
- ✓ tags both rhetorical and linguistic structures
- ✓ brings together composition studies, EFL studies and corpus linguistics

Tagging a feedback sequence

1. First draft, PhD student

Long chain n-3 poly-saturated fatty acids (LC n-3 PUFA), especially EPA and DHA found in...

2. Comments from two PhD students

A: "I think it is better to define them."
B: "Are these known by your readers?"

3. Revised draft (changes underlined)

Long chain n-3 polysaturated fatty acids (LC n-3 PUFA), especially EPA (icosapentaenoic acid) and DHA (docosapentaenoic acid) found in...

Challenges

The **unique design** presents several challenges, for example:

- How do we structure and format the corpus to fit current archiving standards?
- What kind of metadata should be included?
- What kind of interface best serves the researcher and user community?
- To what extent can existing tagging systems be used, and what new tags need to be devised for our material?

The team

Andreas Eriksson, Chalmers, Gothenburg
Damian Finnegan, Asko Kauppinen, Maria Wiktorsson, Anna Wärnsby, Malmö University, Malmö
Peter Withers, MaxPlanck Institute for Psycholinguistics, Nijmegen
Contact: andreas.eriksson@chalmers.se, anna.warnsby@mah.se
Funding: the initial phase of the project is supported by the Crafoord Foundation, Lund, Sweden