

Creating a Corpus using the **WEb-based Learner CORpus MachinE**

Ramon Ziai, Niels Ott

Berlin, October 15, 2009



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Outline

- 1 Project Setup
- 2 Corpus Structure
- 3 Demo of WELCOME
- 4 Discussion and Brainstorming

Goal

- Collecting a corpus consisting of
 - Reading texts
 - Reading comprehension questions
 - Target answers to these questions
 - Answers by students to the questions
 - Assessment of the contents of the student answers
- Language: German

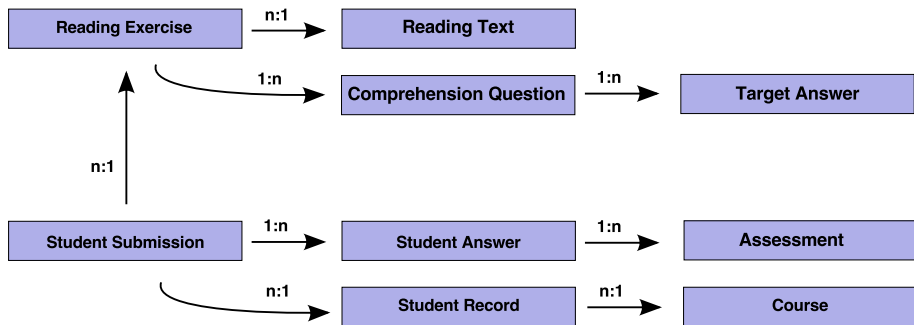
Data Source

- The data are collected from German language classes at two universities in the U.S.
 - Kansas University (Nina Vyatkina)
 - Ohio State University (Kathryn Corl)
- Longitudinal: data are collected during a period of four years.
- . . . from courses of different levels (beginner–intermediate) taught in parallel.

Process of Collecting Data

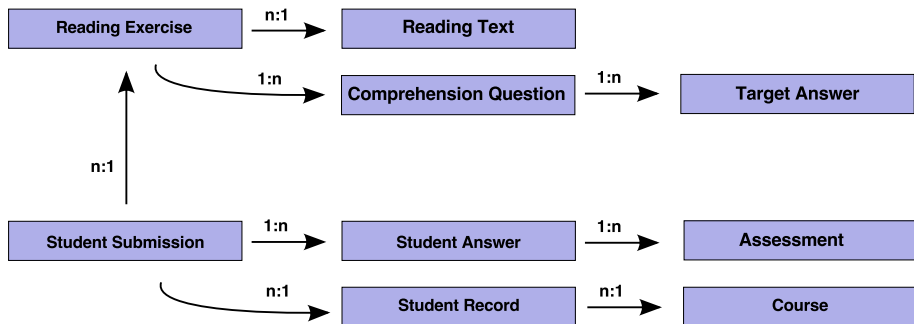
- Data are entered incrementally as structured data by teachers at the two sites.
- Each student answer will be assessed by two annotators independently.
- In addition to the structured data, the exercises and submissions will be scanned in and uploaded.
- This is done using the web-based WELCOME tool developed in our project.

Corpus Structure



- This is a preliminary design.
- Not entirely solved: the student meta data change over time.

Corpus Structure



- This is a preliminary design.
- Not entirely solved: the student meta data change over time.

Demo

`http://delos.sfs.uni-tuebingen.de:8080/welcome`

Thank you!



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

