	Compiling a Task-Based Corpus Detrar Meurers, Nels Os, Ramon Ziai	Outline	Compiling a Task-Based Corpus Detnar Mousers, Niels Ott, Ramon Zail
Compiling a Task-Based Corpus for the Analysis of Learner Language in Context	Project Background and Motivation Comparing meaning in content Collecting data in autheric tasks Compilling a Task-Based Learner Corpus Corpus ingredients	Project Background and Motivation Comparing meaning in context Collecting data in authentic tasks Compiling a Task-Based Learner Corpus	Project Background and Motivation Comparing meaning in contex Collecting data in authentic tasks. Compiling a Task-Bassed Learner Corpus Corpus ingredients
Detmar Meurers, Niels Ott and Ramon Ziai Universität Tübingen, SFB 833	Chair signatures Chairing the Data Chairing the Data Charling the	Corpus ingredients Database structure Obtaining the Data Distributed data collection WELCOME Tool Longitudinality of Meta-Data Content Assessment	Corpus systems of Database servicture Crazining the Data Charlobuse disservation WELCOME Tool Longhadrasily of Mes Crazin Contest Assessment WELCOME Deno Conclusion
Pre-Conference Workshop on Learner Corpora at ALOES 2010 Paris. March 25, 2010	EREFERNINGS UNIVERSITÄT TURINGEN	WELCOME Demo Conclusion	EMERGAN KANAS UNIVERSITÄR TÜRENGEN
Project background	Compiling a Task-Based Corpus Detrar Meurers, Nels Os, Ramon Zial	Collecting data in authentic tasks	Compiling a Task-Based Corpus Detnar Meurers, Nels Ot, Ramon Zail
Project A4 in the SFB 833: Comparing Meaning in Context: Components of a shallow semantic analysis Research question: How can the meaning of sentences and text fragments be analyzed and compared in realistic situations? Realistic situations: Ianguage not necessarily well-formed differences in situative and world knowledge Two challenges: Which linguistic representations can be robustly identified as basis of a computational approximation of meaning? How can the role of the context be integrated?	Project Background and Motivation Extension as service to the Committee of the Compiling of Task-Based Learner Corpus Corpus requires the Compiling of Task-Based Learner Corpus Corpus requires the Compiling of the Corpus Corpus requires the Corpus Corpus requires the Corpus Corpus requires and Corpus requires and Corpus requirements and Corpus reputation and Corpus requirements and Corpus reputation and Corpus reputatio	 We want to make the context explicit by collecting data in the setting of a concrete task. To support evaluation of meaning, focus on tasks using information encoded in language, not world knowledge. In which authentic settings does such data arise? Language in context plays an important role in foreign language teaching (cf., e.g., Ellis 2003). Yet, current learner corpora typically consist of essay data (cf., eg., Granger 2008), so only the essay topic is known; contents often unconstrained and not predictable. Which activities provide more explicit, language-based context? We focus on reading comprehension questions. Compile a corpus with answers to reading 	Project Background and Motivation Company and Section 1997. Company and Section 1997. Company and Section 1997. Se
⇒ Start by collecting data of authentic language in context.	ERERARDAMAN UNIVERSITÄT TERRAGEN 3/14	comprehension questions written by learners of German.	ERIERADEANIA UNIVERSITÄT TURINGEN 4/14

Compiling a Task-Based Corpus Compiling a Corpus ingredients Corpus ingredients Task-Based Cornus Detmar Meurers, Niels Ott Database structure Ramon Ziai 1. Texts asked about in reading comprehension Project Background Project Background and Motivation and Motivation · i.e., the explicit, language-based context Comparing meaning in con Collecting data in authentic Collecting data in authentic 2. Comprehension questions Compiling a Compiling a Task-Based Task-Based Reading Exercise Reading Text Target answers by teachers Learner Cornus Learner Corpus Corpus ingredients Target Answe Student answers Comprehension Question Obtaining the Data Distributed data collection 5. Teacher assessment of student answers Student Submission Student Answe 5.1 binary: correct/incorrect meaning WELCOME Demo 5.2 detailed meaning analysis Student Record Student meta-data: Course Annotato 6.1 age, gender 6.2 native language 6.3 previous exposure to German 6.4 other languages spoken 6.5 ... Compiling a Towards effective distributed data collection Obtaining the Data Task-Based Corpus Task-Based Corpus Detmar Meurers, Niels Ott Project Background Project Background and Motivation Dissociation of source and processing of corpus data: and Motivation Collected in two of the largest German programs in US Collecting data in authentic tasks · Language instructors in the US are the foreign Collecting data in authenti · Our project collaborates with two subcontractors: language teaching experts in touch with the learners. Compiling a Compiling a Task-Based Task-Based · Computational linguists in Germany responsible for Kansas University (Prof. Nina Vvatkina) Learner Corpus Learner Corpus storing and processing the corpus. - Ohio State University (Prof. Kathryn Corl) Corpus ingredients Corpus ingredients ⇒ Distributed data entry, central standardized storage Data is collected Requirements: at four course levels Longitudinality of Meta-Lonoitudinality of Met · over a period of four years. · Entering the data must be straightforward for the language instructors. Conclusion Conclusion Why are we collecting outside of Germany? · Approach must support the complex structure including Controlled context, with a homogeneous group of learners: learner meta-data. typically English native speakers ► How can we meet these requirements? · exposure to German mostly limited to the classroom ⇒ Develop a web-based tool for data collection.

Compiling a Task-Based Corpus Compiling a The WELCOME Tool Longitudinality of Meta-Data Task-Based Cornus Detmar Meurers, Niels Ott, To address the requirements, we developed the Project Background Project Background WEb-based Learner COrpus MachinE (WELCOME). and Motivation Student meta-data change over time It supports distributed data entry by language · e.g., exposure to German instructors and stores all data in a central repository. Compiling a Compiling a ► Ideally, one would collect the meta-data together with Task-Based Task-Based WELCOME behaves similar to a desktop application but Learner Corpus each reading comprehension task. Learner Corpus requires only a web browser and Internet access. Database structure Massive overhead, incompatible with integration of data Obtaining the Data ▶ The interface is collection into regular classes · optimized around the work-flow of language instructors, Compromise: Student meta-data collected once per term. · supports the incremental entry of data resulting in a WELCOME Demo These records are connected via IDs for each student. structured corpus. so we can track each student's development over time. As its back-end, it uses a relational database engine. The reading comprehension answers are stored with a representing and enforcing the complex corpus structure. specific date, supporting more fine-grained tracking of ▶ efficient well-tested development. supports incremental data manipulation and guerving allows concurrent access by multiple users Data can be exported into standard XML formats. Compiling a Content Assessment Content Assessment Task-Based Corpus Task-Based Corpus Detmar Meurers, Niels Ott Example Student answers are assessed by two independent Project Background Project Background and Motivation and Motivation annotators with respect to meaning (not form). Wer glaubt, dass Nikolaus auf Grönland wohnt? 1. The learner answers are independently transcribed Collecting data in authentic 0.6: from the handwritten submissions by each annotator. Compiling a Answer Die Dänen denken dass Nikolaus wohnt auf Grönland. Compiling a Task-Based · Why two annotators? Transcribing handwritten text is an Learner Corpus Learner Corpus Correct Target Answers: interpretation. Corpus ingredients ® Die Dänen glauben, dass Nikolaus auf Grönland wohnt. 2. Binary classification: appropriate vs. inappropriate Obtaining the Data Distributed data-· 'Is the answer given by the student a valid answer to the WELCOME TWI Add alternate correct target... reading comprehension question?' anoitudinality of Meta 3. Fine-grained classification of comparison with target Overall Meaning Assessment: Detailed Meaning Assessment: answers based on Bailey & Meurers (2008) Conclusion Conclusion Correct Correct answer Incorrect For appropriate and inappropriate answers: ..NA.. missing concept, extra concept, blend Correct answer Missing concept Additional answer category for inappropriate answers: Extra concept non-answer Missing and extra concepts · Alternate answers, which are appropriate but differ in contents from target answer, can be added by annotators.



Compiling a Task-Based Corpus Conclusion

Detmar Meurers, Niels Ott, Ramon Ziai

Project Background

Collecting data in authentic

and Motivation

Compiling a

Task-Based

Database structure

Obtaining the Data

Distributed data collection

- We motivated the creation of task-based corpora of
 - German reading comprehension exercises.
 - includes rich structure: context, student data and
- WELCOME tool supports distributed data entry and

 - programs serves as empirical basis for our research on comparing meaning in context. It more generally supports:
 - · SLA research on learner language development linguistic research into language in context, e.g., interaction of syntax and information structure



Compiling a

Task-Based Cornus

Ramon Ziai

Project Background

and Motivation

Compiling a

Task-Based

Learner Corpus

Database structure

Obtaining the Data



- Bailey, S. & D. Meurers (2008). Diagnosing meaning errors in short answers to reading comprehension questions. In J. Tetreault, J. Burstein & R. D. Felice (eds.), Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications, held at ACL 2008, Columbus, Ohio: Association for Computational Linguistics, pp. 107-115, URL http://aclweb.org/anthology-new/W08-0913.
- Ellis, R. (2003). Task-based Language Learning and Teaching. Oxford, UK: Oxford University Press.
- Granger, S. (2008), Learner Corpora in Foreign Language Education, In N. V. Deusen-Scholl & N. H. Hornberger (eds.), Encyclopedia of Language and Education, Volume 4: Second and Foreign Language Education, Springer Science and Business Media, pp. 337-351, 2nd ed.

Compiling a Task-Based Corpus Detmar Meurers, Niels Ott Project Background and Motivation

Comparing meaning in con Collecting data in authentic tasks Compiling a

Learner Corpus Corpus ingredients

Obtaining the Data Distributed data col Longitudinality of Meta-Dat



authentic language data in context.

▶ We are collecting a longitudinal learner corpus of

meta-data, teacher targets and assessment

central, standardized corpus storage.

We will make the tool freely available for research.

► Corpus resulting from our collaboration with US German