

Using NLP Technology in CALL

Cara Greene, Katrina Keogh,
Thomas Koller, Joachim Wagner,
Monica Ward, Josef van Genabith

June 17th 2004



National Centre for Language Technology
School of Computing, Dublin City University



Using NLP Technology in CALL

- Background
- Research methodology
- Activities
 - Plurilingual ICALL System for Romance Languages
 - Artificial Co-Learner
 - ICALL in the Primary School
 - ICALL for Learners with Learning Difficulties
 - ICALL for LCTL
- Summary of research/findings to date



National Centre for Language Technology
School of Computing, Dublin City University



Background of the ICALL Group

- Computational linguists with an interest in CALL
- Six researchers
 - computational linguists
 - software engineers
 - expertise includes
 - general NLP skills, corpus processing
 - CALL, teaching experience
- Interested in different learner types
 - Beginners to advanced, young learners to adults



National Centre for Language Technology
School of Computing, Dublin City University



Research Methodology

- Re-use of existing technologies
 - avoiding “re-inventing the wheel”
- Learning from other ICALL projects
 - avoiding known pitfalls
- Learner-centred design
 - focusing on the needs of the learner
 - taking into account pedagogy and design
 - design for concurrent evaluation



National Centre for Language Technology
School of Computing, Dublin City University



Plurilingual ICALL System

- Target learner
 - advanced speaker of at least one Romance language
 - French, Spanish and Italian supported
 - target language(s): one or two of the other
- Idea
 - leverage the learner’s existing knowledge of already learned Romance language
 - not learning a new language from scratch



National Centre for Language Technology
School of Computing, Dublin City University



Plurilingual ICALL System

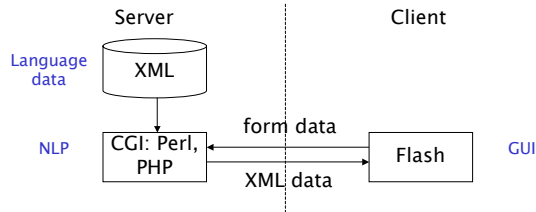
- NLP technologies
 - plurilingual error-sensitive island parser
 - animated grammar presentations
 - use of small, specialised corpora
- ICALL system features
 - ability to select languages of multi-lingual content
 - languages of instruction: English or German



National Centre for Language Technology
School of Computing, Dublin City University



Plurilingual ICALL System



National Centre for Language Technology
School of Computing, Dublin City University



Plurilingual ICALL System

- Re-use of technology
 - error-sensitive island parser for Spanish
 - corpora
- Learn from other projects
 - increasing language production skills (writing)
- Learner-centred
 - explorative learning
 - evaluation platform for continuous assessment



National Centre for Language Technology
School of Computing, Dublin City University



Artificial Co-Learner

- Target learner
 - intermediate to advanced learner of German and English
- Idea
 - exploit inherent limitations of NLP to our advantage
 - the advanced learner "teaches" the artificial co-learner when it makes errors with the L2
 - improve both the human's and computer's L2 knowledge



National Centre for Language Technology
School of Computing, Dublin City University



Artificial Co-Learner

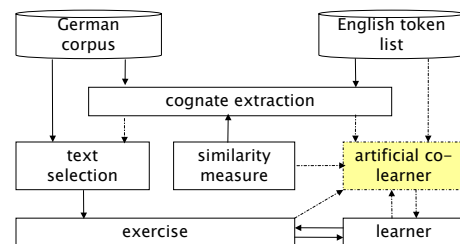
- NLP technologies
 - lemmatisation, POS tagging
 - string similarity measure
 - corpus processing tools
- ICALL system features
 - a tool to automatically create "Cognate and False Friends" learning exercises for the learner



National Centre for Language Technology
School of Computing, Dublin City University



Artificial Co-Learner



National Centre for Language Technology
School of Computing, Dublin City University



Artificial Co-Learner

- Re-use of technology
 - IMS TreeTagger
 - standard string similarity measure
- Design for Evaluation
 - record time spent by learner
 - questionnaire
 - preliminary evaluation with 6 subjects



National Centre for Language Technology
School of Computing, Dublin City University



ICALL in the Primary School

- Two systems: Irish and German
- Target learner
 - 7 – 13 year old (male) pupils in Primary School
 - Target languages:
 - Irish: compulsory (7–13 year olds)
 - German: offered by some schools (10–13 year olds)
- Idea
 - limited L1 knowledge
 - “controlled” L2 knowledge



National Centre for Language Technology
School of Computing, Dublin City University



ICALL in the Primary School: Irish

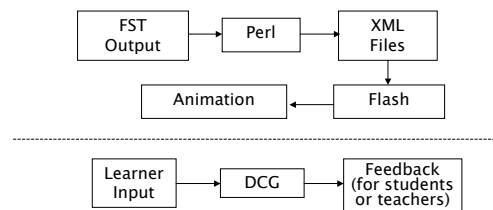
- NLP technologies
 - FST morphology engine for Irish
 - simple, small coverage DCGs
- ICALL systems
 - automatically animated verb conjugations (FST, Perl, XML, Flash)
 - analysis of learner texts (DCGs)



National Centre for Language Technology
School of Computing, Dublin City University



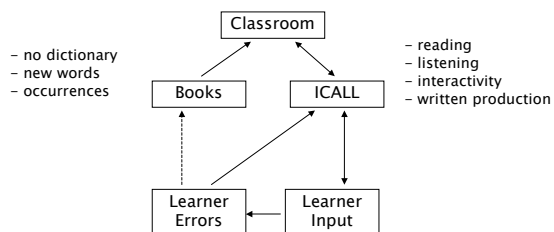
ICALL in the Primary School: Irish



National Centre for Language Technology
School of Computing, Dublin City University



ICALL in the Primary School: Irish



National Centre for Language Technology
School of Computing, Dublin City University



ICALL in the Primary School: German

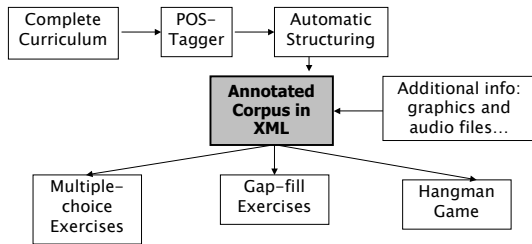
- NLP technologies
 - POS tagger
 - tailored corpus
- ICALL system features
 - annotated XML corpus
 - based on NCCA guidelines for the curriculum
 - enhanced with texts, graphics and audio
 - tools to automatically create exercises



National Centre for Language Technology
School of Computing, Dublin City University



ICALL in the Primary School: German



National Centre for Language Technology
School of Computing, Dublin City University



ICALL in the Primary School

- Re-use of technology
 - FST morphological engine (Úí Dhonnchadha 2002)
 - DCG parser
 - POS tagger (IMS, Schmidt 1994)
 - in-house XML / Flash resources
- Assessment of available & relevant (I)CALL systems
- Learner- (& teacher-) centred approach
 - design for evaluation
 - in line with existing obligatory materials
 - limited L2 knowledge and time to prepare course materials



National Centre for Language Technology
School of Computing, Dublin City University



Conclusion

- Extensive re-use of existing NLP technologies
- Learn from other ICALL projects
- Learner-centred designs
- Design for concurrent evaluation
- NLP is useful not only for CALL for adult and advanced learners, but also for young and ab-initio learners
- Exploit / circumvent limits of NLP



National Centre for Language Technology
School of Computing, Dublin City University



Publications

- K. Keogh, T. Koller, M. Ward, E. Úí Dhonnchadha, & J. van Genabith. 2004. CL for CALL in the Primary School. eLearning for Computational Linguistics and Computational Linguistics for eLearning. International Workshop in Association with COLING 2004, Geneva, Switzerland.
- T. Koller. 2003. Knowledge-based intelligent error feedback in a Spanish ICALL system. In *Proceedings of The 14th Irish Conference on Artificial Intelligence & Cognitive Science*. Dublin: Trinity College, 117-121.
- T. Koller. 2004. Entwicklung eines multilingualen ICALL-Systems für Französisch, Italienisch und Spanisch. To be published in: H.G. Klein / D. Rutke: *Neuere Forschungen zur europäischen Interkomprehension*. Aachen: Editiones EuroCom (vol. 21).
- J. Wagner. (to appear). A false friend exercise with authentic material retrieved from a corpus. In *Proceedings of INSTIL / ICALL 2004*, Venice, Italy



National Centre for Language Technology
School of Computing, Dublin City University



References

- E. Úí Dhonnchadha. 2002. An Analyser and Generator for Irish Inflectional Morphology Using Finite-State Transducers. MSc Thesis, Dublin City University, Ireland
- A. McEnery and M.P. Oakes. 1996. Sentence and Word Alignment in the CRATER Project. In J. Thomas and M. Short (eds) *Using Corpora for Language Research*, Longman, pp 211-231
- Flash. <http://www.macromedia.com/software/flash/>
- H. Schmidt. 1994. Probabilistic Part-of-Speech Tagging using Decision Trees. <http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger1.pdf>
- XML. <http://www.w3.org/XML/>



National Centre for Language Technology
School of Computing, Dublin City University



Thank You!

Discussion



National Centre for Language Technology
School of Computing, Dublin City University

