

Dr. Carolyn P. Rose
cprose@cs.cmu.edu
CARNEGIE MELLON UNIVERSITY
Language Technologies Institute/ Human-Computer Interaction Institute
5000 Forbes Avenue
Pittsburgh, PA
15213
Phone: 412-268-7130
Fax: 412-268-6298

Website URL: <http://www.cs.cmu.edu/~cprose>
Project Title: A shared Resource for Robust Semantic Interpretation for Both
Linguists and Non-linguists
ONR Award No: N000140510043
Organization Award No: 12599-1-1140108
Reporting Period: June 1, 2006 - May 31, 2007
Award Period: November 8, 2004 - September 30, 2007

Scientific and Technical Objectives

Our goals are still the same as those that were in our proposal. Specifically:

The overall goal of the TFlex project is to make language technology easier to use in tutorial dialogue and learning research. Specifically, we need the following

1. Tools for fast and robust deep parsing of natural language, supporting tutorial dialogue and computer-assisted learning
2. Lexical resources to be used in parsing, and tools to extend their coverage to new domains
3. Tools that make language technology accessible to domain and educational experts.

Approach

We have developed a set of language processing tools supporting different

aspects of educational technologies related to language interpretation (tutorial dialogue systems, computer-supporting collaborative learning, learning sciences research).

The tools for extending linguistic resources that we are developing will allow researchers and educators to bring deeper linguistic features into classification tasks, improving classification accuracy and helping with data analysis. They will also support development of lexicons for deep parsing, which are necessary for understanding more complex language. Better language understanding can then be used to help build better tutorial dialogue systems, and to improve automatic assessment tools.

Concise Accomplishments

First Significant Accomplishment: A set of tools for educational researchers for analysis of collaborative learning and development of tutorial dialogue interfaces:

The CMU team developed a set of tools for educational researchers. This research takes a user-centered design approach to making natural language technology accessible to non-linguists by understanding their naïve conceptions of language and creating environments where natural language technology can be used in a way that matches the way non-linguists think and work.

Second Significant Accomplishment: A Robust Deep Parser For Dialogue:

In tutorial situations, student input is often not well-formed, and speed and robustness are major issues in interpreting the student's utterances. The collaboration between Edinburgh and CMU resulted in building the TFLex robust parser (Dzikovska & Rosé, 2005; Dzikovska & Rosé, 2006) which integrates the TRIPS deep interpretation framework (Dzikovska, 2004) with the LCFlex robust parsing architecture (Rosé, 2000; Rosé & Lavie, 2001; Rosé et al., 2002).

Third Significant Accomplishment: Developing Lexical Resources for Deep Parsing

The Edinburgh team developed tools to improve coverage of deep parsers. We developed methods to extract lexical information from existing lexical resources, and put it in a form useful for parsing.

Expanded Accomplishments

First Significant Accomplishment: A set of tools for educational researchers for analysis of collaborative learning interactions and development of tutorial dialogue interfaces:

The CMU team developed a set of tools for educational researchers. This research takes a user-centered design approach to making natural language technology accessible to non-linguists by understanding their naïve conceptions of language (Rosé, Pai, & Arguello, 2005) and creating environments where natural language technology can be used in a way that matches the way non-linguists think and work.

1. TuTalk tutorial dialogue authoring environment (Gweon et al., 2005), discussed in connection with the TuTalk dialogue engine (Jordan et al., to appear). Note that the TuTalk dialogue engine was developed under separate funding. The TuTalk authoring environment includes:

A. The InfoMagnets corpus organization tools (Arguello & Rosé, 2006a; Arguello & Rosé, 2006b; Arguello & Rosé, 2006c)

B. The TagHelper language understanding toolkit (Rosé et al., under review; Wang et al., to appear; Donmez et al., 2005) makes it easy for non-linguists to extract linguistic features from texts that can be used for processing student answers

C. A GUI interface for authoring, previewing, and testing dialogue agents

2. A toolset for supporting learning sciences research beyond tutorial dialogue systems – a second application of TagHelper tools. TagHelper tools has been used for the following purposes:

A. facilitating process analyses that allow us to understand the impact of educational interventions on behavior and learning (Wang et al., to appear-c; Rosé et al., under review)

B. filtering conversational interactions in form of chat, email, and newsgroup interaction for the purpose of understanding student explanations (Rosé et al., Under Review; Rosé & VanLehn, 2005)

D. triggering interventions in a collaborative learning setting (Kumar et al., to appear; Wang et al., to appear)

D. on-line assessment of learning (Joshi & Rosé, submitted) and collaboration

processes in an instructional context (Wang et al., to appear-b; Rosé et al., to appear).

The tools have been used significantly by tutorial systems designers and educational researchers:

1. The TuTalk and TagHelper tools packages were first piloted in summer of 2006 at the Pittsburgh Science of Learning Center summer school with approximately 30 participants.
2. TagHelper tools has been demonstrated at an International Conference of the Learning Sciences workshop on adaptive collaborative learning support as well as being featured in a plenary talk and two invited workshop talks at the 2007 Kaleidoscope Alpine CSCL Rendez Vous in Villars, Switzerland.
3. Improved versions will be released at the 2007 Pittsburgh Science of Learning Center and will each be the topic of a separate tutorial offered at AI in Education 2007. We also plan to run a workshop on analysis tools for computer supported collaborative learning at the International Conference of the Learning Sciences in summer of 2008.
4. TagHelper tools has been used as a teaching tool in three different graduate courses at Carnegie Mellon University including Machine Learning in Practice (Language Technologies Institute), Research Methods in the Learning Sciences (Psychology), and Computer-Supported Collaborative Learning (Human-Computer Interaction Institute). Instructors at other universities, such as Kwangsu Cho at the University of Missouri and Stefan Trausan-Matu at Politechnica University of Bucharest, also plan to use it in their courses.
5. Researchers in learning sciences, political science, social psychology, and human-computer interaction are using TagHelper tools to support their discourse analysis work. We have 39 users on our active users mailing list.

Second Significant Accomplishment: A Robust Deep Parser For Dialogue:

In tutorial situations, student input is often not well-formed, and speed and robustness are major issues in interpreting the student's utterances. The collaboration between Edinburgh and CMU resulted in building the TFLex robust parser (Dzikovska & Rosé, 2005; Dzikovska & Rosé, 2006) which integrates the TRIPS deep interpretation framework (Dzikovska, 2004) with the LCFlex robust parsing architecture (Rosé, 2000; Rosé & Lavie, 2001; Rosé et al., 2002). Compared to the TRIPS parser, TFLex provides significantly faster

processing for long sentences, with a higher probability of finding a spanning parse.

Third Significant Accomplishment: Developing Lexical Resources for Deep Parsing

Currently, tutorial dialogue systems only allow very limited forms of student input, and all data annotation is done entirely by hand. Deep parsing can facilitate the detailed analysis of student input necessary in automated assessment, both for tutorial dialogue, and for analysing student essays. In addition, linguistic features can improve classification accuracy (Rosé & VanLehn, 2005). However, deep parsers are often difficult to extend to new domains because of their limited lexical coverage, making them difficult to use even for trained linguists. The Edinburgh team developed tools to improve coverage of deep parsers. We developed methods to extract lexical information from existing lexical resources (used in question answering and information extraction), and put it in a form useful for parsing. More details are included in the Edinburgh report.

Uploaded Files:

[FLAIRS05RoseC.pdf](#)

[HLT-Demo06.pdf](#)

[IJCSCL-S-07-00020.pdf](#)

[Museli_Final_v2.pdf](#)

[scanalu04.pdf](#)

[SLaTE07-JoshiM.pdf](#)

[TuTalkPaper-Camera.pdf](#)

Work Plan

1. Expanding the scope of tools for non-linguists in order to increase the potential impact. Our work to date shows promise that the technology we have been developing is useful not only in the context of tutorial dialogue interactions, but also for on-line assessment of learning processes and productivity as well as for triggering contextually appropriate interventions in computer supported collaborative learning settings. In the renewal grant, the CMU team would like to focus significant effort on this expanded scope, moving from a focus on building a tool set for content developers to building applications to support instruction either indirectly through supporting instructors or directly through triggering interventions that support learning interactions.

Note: both of these efforts produce systems that can be easily ported to new domains because they build on the tool sets we have developed in our work so far.

A. Build a system to support group learning facilitators, leveraging preliminary work (Joshi & Rosé, submitted; Rosé et al., to appear). This includes design and implementation of the reporting interface that instructors will interact with as well as basic research on assessment from conversational data.

B. Further develop our architecture for adaptive collaborative learning support (Kumar et al., 2007; Kumar et al., submitted), which uses conversational agents to support collaboration. This involves further investigating a state-based approach to modeling collaborative learning interactions and development of more socially sensitive conversational agents to address problems we have seen now in three studies using dialogue agents in a collaborative learning context (Kumar et al., 2007; Wang et al., 2007; Kumar et al., submitted). Although this architecture was first developed in the context of the CycleTalk project, its continued development is not germane to that project, thus it will not continue to be developed under that funding.

2. Integrate the developed resources in the Beetle tutorial system, to improve language interpretation in order to better support students in deep explanation during tutorial dialogue interactions. This will be a major contribution of the Edinburgh team. If successful, more advanced conversational capabilities enabled by this technology can eventually be migrated to the TuTalk authoring environment (see Step 3 below).

A. Extend the algorithms to extract lexical entries for nouns and adjectives

B. Merge the lexicon extracted from FrameNet with the TRIPS lexicon for use in the Beetle system

C. Evaluate improvements in parsing and interpretation quality (in the Beetle system) when using the extended lexicon

3. Integrate the developed resources into the toolset for non-linguists, to improve their effectiveness and usability

A. Integrate the TRIPS/TFlex parser with the TagHelper tool, and evaluate the impact of deeper linguistic features on classification accuracy (CMU team).

B. Extend the TuTalk tools to take advantage of the more detailed analyses output by the parser (CMU team).

C. Add mappings from the FrameNet and TRIPS lexicons to a robust parser for text, to support analyzing essays (Edinburgh team).

4. Develop tools for quickly adding domain-specific vocabulary when a new domain needs to be analyzed (Edinburgh team).

Problems/Issues

We did not receive our funds for this year, so in order to make progress on this project, we needed to borrow resources from the CycleTalk project, which slowed down progress on CycleTalk. That will be addressed in the CycleTalk report.

Peer-Reviewed Journal Articles

No journal articles reported.

Books or Book Chapters

No book or book chapters reported.

Technical Reports (Non-refereed Publications)

Rosé, C. P., Wang, Y.C., Cui, Y., Arguello, J., Fischer, F., Weinberger, A., Stegmann, K. (Under Review). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning, submitted to the International Journal of Computer Supported Collaborative Learning

Stegmann, K., Weinberger, A., Fischer, F., & Rosé, C. P. (2006). Automatische Analyse nat,rllich-sprachlicher Daten aus Onlinediskussionen [Automatic corpus analysis of natural language data of online discussions]. Paper presented at the 68th Tagung der Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF, Working Group for Empirical Educational Research) Munich, Germany.

Arguello, J. & Rosé, C. P. (2006). InfoMagnets: Making Sense of Corpus Data, Companion Proceedings for the North American Chapter of the Association for Computational Linguistics (NAACL '06). (one of three demos selected for presentation in a plenary session)

Abstracts/Presentations/Posters/Conference Proceedings

Conference: Joshi, M. & Rosé, C. P. (to appear). Using Transactivity in Conversation Summarization in Educational Dialog. Proceedings of the SLATE Workshop on Speech and Language Technology in Education

Poster: Wang, Y. C., Joshi, M., & Rosé, C. P. (2007b). A Feature Based Approach for Leveraging Context for Classifying Newsgroup Style Discussion Segments, Proceedings of the Association for Computational Linguistics (poster).

Poster: Wang, Y., Rosé, C. P., Joshi, M., Fischer, F., Weinberger, A., Stegmann, K. (to appear). Context Based Classification for Automatic Collaborative Learning Process Analysis, Proceedings of Artificial Intelligence in Education (poster).

Conference: Arguello, J. & Rosé, C. P. (2006). Topic Segmentation of Dialogue, Proceedings of the NAACL Workshop on Analyzing Conversations in Text and Speech.

Conference: Dzikovska, M. & Rosé, C. P. (2006). Backbone Extraction and Pruning for Speeding Up a Deep Parser for Dialogue Systems, Proceedings of the 3rd International Workshop on Scalable Natural Language Processing (ScaNaLU).

Conference: Arguello, J. & Rosé, C. P. (2006). Museli: A Multi-source Evidence Integration Approach to Topic Segmentation of Spontaneous Dialogue, Proceedings of the North American Chapter of the Association for Computational Linguistics (short paper)

Poster: Dzikovska, M. & Rosé, C. P. (2005). TFLEX: Making Deep Parsing Practical with Strategic Pruning, Proceedings of the International Workshop on Parsing Technologies (poster)

Conference: Gweon, G., Arguello, J., Pai, C., Carey, R., Zaiss, Z., Rosé, C. P. (2005). Towards a Prototyping Tool for Behavior Oriented Authoring of Conversational Interfaces, Proceedings of the ACL Workshop on Educational Applications of NLP.

Conference: Rosé, C. P., Pai, C., Arguello, J. (2005). Enabling Non-linguists to Author Conversational Interfaces Easily, Proceedings of FLAIRS 05.

Awards/Honors/Invention Disclosure

Language Technologies for Supporting Productive Collaborative Learning Interactions for Science and Engineering Education, Keynote talk, Technology-integrated Science and Engineering Education Workshop (TechSEE-II), National Taiwan Normal University, May 2007

Towards Triggering Adaptive Collaboration Support Using Automatic Interaction Analysis, Community Day Plenary talk, Kaleidoscope CSCL Rendez Vous, January 2007

Towards Adaptive Collaboration Support, Workshop on Computer Supported Collaboration Scripts, Keynote talk, Kaleidoscope CSCL Rendez Vous, January 2007

TagHelper: Computer Support for Applying Coding Schemes, Keynote talk, Workshop on Computer Based Analysis and Visualization of Collaborative Learning Activities, Kaleidoscope CSCL Rendez Vous, January 2007

Towards Adaptive Support for On-line Learning, Keynote talk, Technology-integrated Science and Engineering Education Workshop (TechSEE), National Taiwan Normal University, May 2006

Making Authoring of Conversational Interfaces Accessible, Featured Talk, Workshop on Authoring Tools for Advanced Learning Systems with Standards (organized by Arthur Graesser, The Advanced Distributed Learning Workforce Co-Lab at the University of Memphis), November 2005

Patents Submitted

No patents submitted reported.

Patents Issued

No patents issued reported.

Technology Transfer

TagHelper tools is licensed under the General Public License (GPL). We are working out a licensing arrangement for the TuTalk authoring environment. Worth Publishers is interested in using it for developing some interactive materials to distribute with their textbooks.

ONR Database Statistics

Use of Human Subjects	Yes
DoD Personnel Used	
Use of Animals	No
Animals Used	
Use of Recombinant DNA	No
Degree(s) Granted	1

PI/CoPI Information				
0 PI/CoPI Minority Women**				
1 PI/CoPI Non-Minority Women	Women	0	1	1
0 PI/CoPI Minority Men**	Men	0	0	0
0 PI/CoPI Non-Minority Men	Total	0	1	1

Post Doctoral Information				
0 Post Doctoral Minority Women**				
0 Post Doctoral Non-Minority Women	Women	0	0	0

0 Post Doctoral Minority Men**	Men	0	0	0
0 Post Doctoral Non-Minority Men	Total	0	0	0

Grad Students Information				
0 Grad Students Minority Women**				
0 Grad Students Non-Minority Women				
0 Grad Students Minority Men**				
1 Grad Students Non-Minority Men				
		Minority Non-Minority Total		
	Women	0	0	0
	Men	0	1	1
	Total	0	1	1

Undergrad Students Information				
0 Undergrad Students Minority Women**				
0 Undergrad Students Non-Minority Women				
0 Undergrad Students Minority Men**				
0 Undergrad Students Non-Minority Men				
		Minority Non-Minority Total		
	Women	0	0	0
	Men	0	0	0
	Total	0	0	0

Publication Totals	
Total Number of Peer-Reviewed Journal Articles:	0
Total Number of Books or Chapters:	0
Total Number of Technical Reports:	3
Total Number of Abstracts/Presentations/Posters/Conference Proceedings:	9
Total Number of Patents Issued:	0

Total Number of Patents Pending:

0

*** Under-represented or minority groups include Blacks, Hispanics, and Native Americans. Asians are not considered an under-represented or minority group in science and engineering.*

**** Supported at least 25% this year on contract/grant.*

Other Sponsored Work

TITLE	SPONSOR	FUNDING	START DATE	END DATE
Pittsburgh Science of Learning Center	National Science Foundation	500K	September 1, 2004	February 29, 2008

Foreign Collaborations

No foreign collaborations reported.

Uploads

[FLAIRS05RoseC.pdf](#)

[HLT-Demo06.pdf](#)

[IJCSCSCL-S-07-00020.pdf](#)

[Museli_Final_v2.pdf](#)

[scanalu04.pdf](#)

[SLaTE07-JoshiM.pdf](#)

[TuTalkPaper-Camera.pdf](#)

[Required_Viewgraph_CycleTalk07.ppt](#)