

# Preface

The general purpose of this symposium is to strengthen collaboration between researchers and users of language learning tools, by fostering common applied and theoretical interests in the communities of two independent conferences held at the University of Maryland: the 38<sup>th</sup> annual meeting of the Association for Computational Linguistics (ACL-99) and the biennial meeting of the International Association of Language Learning Technologies (IALL-99). The call for participation solicited presentations broadly: language assessment; software for first and second language acquisition; general discussions of problems and solutions in evaluating systems, software, and people learning language; integrating technology and foreign language pedagogy; user studies; the relation between technology and language learning, or linguistic theory and tool building; discussions of what is feasible and/or desirable in language learning technology; computer adaptive testing and speech recognition in language assessment and placement; software and tool demonstrations; formulation of discussion questions for panels relating to any of the above.

Sixteen papers were submitted, and 10 accepted for presentation based on blind reviews. The reviewers are listed on the program page, and hereby heartily thanked for their willing and constructive participation. I would also like to thank the the IALL Conference Chair Chris Higgins, ACL Business Administrator Priscilla Rasmussen, and the ACL workshop coordinators, Susan Armstrong and Derek Walker for their help making the mechanics of the symposium run smoothly, despite the challenges of inter-conference communication. For inspiration, thanks goes again to the review committee, with special gratitude for the input of Dorry Kenyon and Valerie Malabonga (whose demonstration at the University of Maryland inspired the symposium idea), Carol Van Ess-Dykema and Chris Higgins (again).

The presentations represent in large part research combining work in natural language processing (NLP) and language learning, including tool development for teaching and assessing language learning. A smaller number of presentations dealt with evaluation of machine language acquisition, as in machine translation applications. The papers are organized into three sessions, based on the type of language they take as input. Four papers use or evaluate **oral language**; two address **multilingual and multimodal issues**; and four deal with **written language**. The following questions characterize the main concerns of the workshop, cross-cutting the modality of the language.

## **What does it mean for a human or machine to ‘know’ a language?**

Successful language learners know what to say or write, when, and for what purpose, to say or write it, and how to manage miscommunication. A learner may have memorized specific rules for various constructions, but his/her overt knowledge is only a small part of the grammar that has been internalized.

Successful language systems also analyze and generate language appropriately, with robust strategies for recovering from failure. However, automated systems require explicit models of language. Additional modifications may be required for rendering this knowledge in a computationally feasible system. What is known about the human language and how it is learned

can be used in system creation, but only if it is articulated in sufficient detail. Conversely, modeling language for NLP systems may potentiate learning about human language. In the oral language session, Tomokiyo and Burger outline a project to collect and model the speech of English language learners whose native language is Japanese. The resulting data could be applied to model learner language in an oral evaluation task, and perhaps generalize to other non-standard speech.

Michaud and McCoy model written proficiency to enhance acquisition of English writing by native users of American Sign Language. Their design provides input to an assessment of writing, as well as error feedback at an appropriate level. Heift and McFetridge also use a student model for assessment and error feedback on constructions the student is capable of acquiring, given current competence. Their paper focuses on reducing parse ambiguity in a tutoring system for English speakers learning German.

### **What is involved in language assessment/evaluation?**

Both learners and systems can be evaluated in terms of competence on various tasks (including oral and written production and comprehension) and statistical standards developed for rating competence. Some standards are better developed in certain areas than others: parsing standards, for example, are much better articulated in the NLP literature than numerical assessments of grammatical competence for language learners. Conversely, pronunciation assessment in learners (and what possible errors to look for, based on contrastive analysis between mother and target languages) is more well-developed than standards for speech synthesis and recognition at the level of phonemes and words.

Robust NLP systems can be evaluated on the explicit models of languages they require. The coverage of the system can be measured by the size of the components: how big is the lexicon, how many constructions is the grammar designed to cover, how general is the discourse strategy encoded in the system, etc. . These components are not amenable to inspection in human language assessment. In addition, task-based evaluation may be conducted without regard for the affective components, such as fatigue, cultural norms, politeness, etc. In contrast, human language assessment is inseparable from such considerations. Even if the teaching/assessment process could be effectively modeled (passing some version of a Turing test), psychometric evaluation may reveal variation due solely to whether the student is interacting with a computer or a person.

Both oral and text assessment task may be subdivided into test creation, administration, and scoring. With respect to oral language, Fairfield outlines a method using Rosetta Stone™ software for automatic test creation and administration. Scoring is implicit: students assess their own speech, using automatically generated feedback. Malabonga and Kenyon describe an oral assessment for Spanish language learners, with systems under development for Arabic and Mandarin Chinese. The test items were created and will be assessed manually; however, each test is compiled interactively in collaboration with the student, a design intended to reduce the examinee's affective reaction to a computer-administered exam. Levow and Olsen discuss what is involved in automating the scoring of such a test, further dividing evaluation into a process (how it is done) and a result (assigning a particular score). In principle, computer-generated speech could be assessed using one or more of the paradigms above. Such an assessment could provide a standard of evaluation for speech generation.

Decrozant and Voss propose using the same standards and system for evaluating human and machine-generated text. They evaluate English-speaking learners of French and an English-French machine translation system on spatial expressions, based on contrastive analysis of the languages. Lenci, Montemagni, Pirrelli and Soria similarly suggest that a single annotation scheme can be used to permit comparison and evaluation of parsers that take different languages as input, and in either oral or written form.

Fairon describes a European Union project to create and administer written language tests automatically from manually crafted test items and language models. The paper focuses on native French; Italian and German test projects are also under development. Burstein and Chodorow discuss automatic assessment of English tests that have been manually created and administered. They outline a method for scoring the exams electronically, identifying lexical, syntactic and discourse features.

It is our hope that this symposium provides answers to what it means for both people and systems to 'know' and 'evaluate' language, by highlighting the benefits in cross-fertilization of the language learning and natural language processing fields.

Mari Broman Olsen, Chair  
May 1999  
University of Maryland, USA