Tactical Language Training System: An Interim Report

W. Lewis Johnson[1], Carole Beal[1], Anna Fowles-Winkler[2], Ursula Lauper[2], Stacy Marsella[1], Shrikanth Narayanan[3], Dimitra Papachristou[1], and Hannes Vilhjálmsson[1]

[1] Center for Advanced Research in Technology for Education (CARTE), USC / Information Sciences Institute, 4676 Admiralty Way, Marina del Rey, CA 90292 USA
{Johnson, CBeal, Marsella, Dimitrap, Hannes}@isi.edu
[2] Micro Analysis & Design, 4949 Pearl East Circle, Suite 300, Boulder, CO 80301 USA
{AWinkler, ULauper}@maad.com
[3] Speech Analysis and Interpretation Laboratory, 3740 McClintock Avenue, Room EEB 430 Los Angeles, CA 90089 2564
Shri@sipi.usc.edu

Abstract. Tactical Language Training System helps learners acquire basic communicative skills in foreign languages and cultures. Learners practice their communication skills in a simulated village, where they must develop rapport with the local people, who in turn will help them accomplish missions such as post-war reconstruction. Each learner is accompanied by a virtual aide who can provide assistance and guidance if needed, tailored to each learner's individual skills. The aide can also act as a virtual tutor as part of an intelligent tutoring system, giving the learners feedback on their performance. Learners communicate via a multimodal interface, which permits them to speak and choose gestures on behalf of their character in the simulation. The system employs video game technologies and design techniques, in order to motivate and engage learners. A version for Levantine Arabic has been developed, and versions for other languages are in the process of being developed.

1 Introduction

The Tactical Language Training System helps learners acquire communicative competence in spoken Arabic and other languages. An intelligent agent coaches the learners through lessons, using innovative speech recognition technology to assess their mastery and provide tailored assistance. Learners then practice particular missions in an interactive story environment, where they speak and choose appropriate gestures in simulated social situations populated with autonomous, animated characters. We aim to provide effective language training both to high-aptitude language learners and to learners with low confidence in their language abilities. We hypothesize that such a learning environment will be more engaging and motivating than traditional language instruction and yield rapid skill acquisition and greater learner self-confidence.

2 Motivations

Current foreign language instruction is heavily oriented toward a small number of common languages. For example, in the United States, approximately ninety-one percent of Americans who study foreign languages in schools, colleges, and universities choose Spanish, French, German, or Italian, while very few choose such commonly spoken languages such as Chinese, Arabic, or Russian [18]. Arabic, the sixth most widely spoken language in the world, accounts for less than 1% of US college foreign language enrollment [17]. Moreover, many such courses can be very time consuming, because learners often must cope with unfamiliar writing systems as well as differing cultural norms. This can be a significant barrier for students who want to acquire basic communication skills so that they can function effectively overseas.

The Tactical Language Training System (TLTS) provides integrated training in foreign spoken language and culture. It employs a task-based approach, where the learner acquires the skills needed to accomplish particular communicative tasks [4]. It focuses on authentic tasks of particular relevance to the learners, involving social interactions with (simulated) native speakers. Written language is omitted, to emphasize basic spoken communication. Vocabulary is limited to what is required for specific situations, and is gradually expanded through a series of increasingly challenging situations that comprise a story arc or narrative. Grammar is introduced only as needed to enable learners to generate and understand a sufficient variety of utterances to cope with novel situations. Nonverbal gestures (both "dos" and "don'ts") are introduced, as are cultural norms of etiquette and politeness, to help learners accomplish the social interaction tasks successfully. We are developing a toolkit to support the rapid creation of new task-oriented language learning environments, thus making it easier to support less commonly taught languages. A preliminary version of a training system has been developed for Levantine Arabic, and a new version for Iraqi Arabic is under development.

Although naturalistic task-oriented conversation has the advantage of encouraging learning by doing, such conversations by themselves do not ensure efficient learning [4], [14]. Learners also benefit from form feedback (i.e., corrective feedback on the form of their utterances) when they make mistakes. But since criticism can be embarrassing and face-threatening [2], native speakers may avoid criticizing learner speech in social situations. Language instructors are more willing to critique learner language; however the language classroom is an artificial environment that easily loses the motivational benefits of authentic task-oriented dialog. The TLTS addresses this problem by providing learners with two closely coupled learning environments with distinct interactional characteristics. The Mission Skill Builder (MSB) incorporates a pedagogical agent that provides continual form feedback. The Mission Practice Environment (MPE) provides authentic practice in social situations, accompanied by an aide character who can offer help if needed. This approach combines task orientation, form feedback, and scaffolding to maximize learning efficiency and effectiveness. The MSB builds on previous work with socially intelligent pedagogical agents [10], [11], while the MPE build on work on interactive pedagogical dramas [15].

The Mission Practice Environment is built using computer game technology, and exploits game design techniques, in order to promote learner engagement and motivation. Although there is significant interest in the potential of game technology to promote learning [6], there are some important outstanding questions about how to exploit this potential. One is *transfer* – how does game play result in the acquisition of skills that transfer outside of the game? Another is how best to exploit *narrative structure* to promote learning? Narrative structure can make learning experiences more engaging and meaningful, but can also discourage learners from engaging in learning activities such as exploration, study, and practice that do not fit into the story line. By combining learning experiences with varying amounts of narrative structure, and by evaluating transfer to real-world communication, we hope to develop a deeper understanding of these issues.

The TLTS builds on ideas developed in previous systems involving microworlds (e.g., FLUENT, MILT) [7],[9], conversation games (e.g., Herr Kommissar) [3], speech pronunciation analysis [23], learner modeling, simulated encounters with virtual characters (e.g., Subarashii, Virtual Conversations, MRE) [1], [8], [20]. It extends this work by providing rich form feedback, by separating game interaction from form feedback, and by supporting a wide range of spoken learner inputs, in an implementation that is robust and efficient enough for ongoing testing and use on commodity computers. The use of speech recognition for tutoring purposes is particularly challenging and innovative,since speech recognition algorithms tend not to be very reliable on learner speech.

3 Example

The following scenario illustrates how the TLTS is used. To appreciate the learner's perspective, imagine that you are a member of an Army Special Forces unit assigned to conduct a civil affairs mission in Lebanon.[1] Your unit will need to enter a village, establish rapport with the people, make contact with the local official in charge, and help carry out post-war reconstruction. To prepare for your mission, you go into the Mission Skill Builder and practice your communication skills, as shown in Figure 1. Here, for example, you learn a common greeting in Lebanese Arabic, "marHaba." You practice saying "marHaba" into your headset microphone. Your speech is automatically analyzed for errors, and your virtual tutor, Nadiim, gives you immediate feedback. If you mispronounce the pharyngeal /H/ sound, as native English speakers commonly do, you receive focused, supportive feedback. Meanwhile, a learner model keeps track of the phrases and skills you have mastered. When you feel that you are ready to give it a try, you enter the Mission Practice Environment. Your character in the game, together with a non-player character acting as your aide, enters the village. You enter a café, and start a conversation with a man in the café, as shown

[1] Lebanon was initially chosen because Lebanese native speakers and speech corpora are widely available. This scenario is typical of civil affairs operations worldwide, and does not reflect actual or planned US military activities in Lebanon.

in Figure 2 (left). You speak for your character into your microphone, while choosing appropriate nonverbal gestures. In this case you choose a respectful gesture, and your interlocutor, Ahmed, responds in kind. If you encounter difficulties, your aide can help you, as shown in Figure 2 (right). The aide has access to your learner model, and therefore knows what Arabic phrases you have mastered. If you had not yet mastered Arabic introductions the aide would provide you with a specific phrase to try. You can then go back to the Skill Builder and practice further.

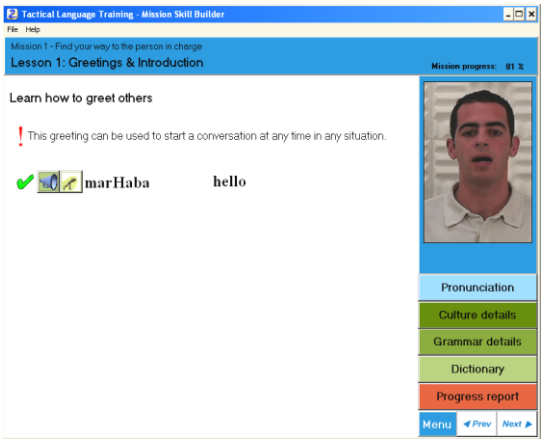

Fig. 1. A coaching section in the Mission Skill Builder

Fig. 2. Greeting a Lebanese man in a café

4 Overall System Architecture

The TLTS architecture must support several important internal requirements. A Learner Model supports run-time queries and updates by both the Skill Builder and the Practice Environment. Learners need to be able to switch back and forth easily

between the Skill Builder and the Practice Environment, as they prefer. The system must support rapid authoring of new content by teams of content experts and game developers. The system must also be flexible enough to support modular testing and integration with the DARWARS architecture, which is intended to provide any-time, individualized cognitive training to military personnel. Given these requirements, a distributed architecture makes sense (see Figure 3). Modules interact using content-based messaging, currently implemented using the Elvin messaging service.

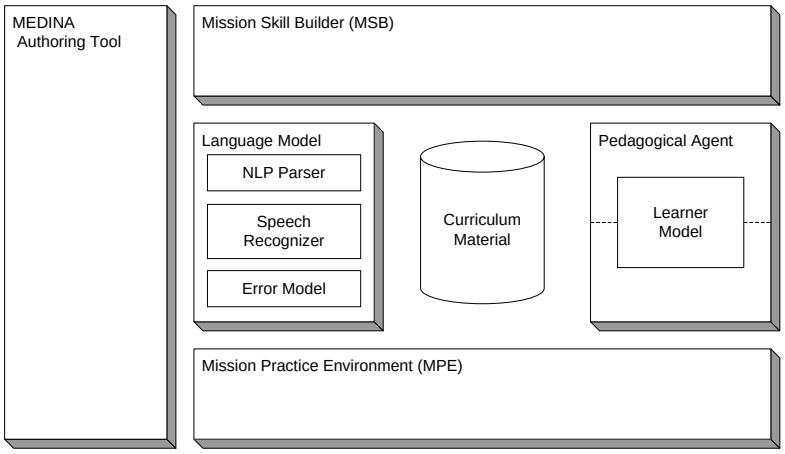

Fig. 3. The overall TLTS architecture

The Pedagogical Agent monitors learner performance, and uses performance data both to track the learner's progress in mastering skills and to decide what type of feedback to give to the learner. The learner's skill profile is recorded in a Learner Model, which is available as a common resource, and implemented as a set of inference rules and dynamically updated tables in an SQL database. The learner model keeps a record of the number of successful and unsuccessful attempts for each action over the series of sessions, as well as the type of error that occurred when the learner is unsuccessful. This information is used to estimate the learner's mastery of each vocabulary item and communicative skill, and to determine what kind of feedback is most appropriate to give to the learner in a given instance. When a learner logs into either the Skill Builder or the Practice Environment, his/her session is immediately associated with a particular profile in the learner model. Learners can review summary reports of their progress, and in the completed system instructors at remote locations will be able to do so as well.

To maintain consistency in the language material, such as models of pronunciation, vocabulary and phrase construction, a single Language Model serves as an interface to the language curriculum. The Language Model includes a speech recognizer that both applications can use, a Natural Language Parser that can annotate phrases with structural information and refer to relevant grammatical explanations and an Error Model which detects and analyzes syntactic and phonological mistakes.

While the Language Model can be thought of as a view of and a tool to work with the language data, the data itself is stored in a separate Curriculum Materials database. This database contains all missions, lessons and exercises that have been constructed, in a flexible Extensible Markup Language (XML) format, with links to media such as sound clips and video clips. It includes exercises that are organized in a recommended sequence, and tutorial tactics that are employed opportunistically by the pedagogical agent in response to learner actions. The database is the focus of the authoring activity. Entries can be validated using the tools of the Language Model. The Medina authoring tool (currently under development) consolidates this process into a single interface where people with different authoring roles can view and edit different views of the curriculum material while overall consistency is ensured.

Since speech is the primary input modality of the TLTS, robustness and reliability of speech processing are of paramount concern. The variability of learner language makes robustness difficult to achieve. Most commercial automated speech recognition (ASR) systems are not designed for learner language [13], and commercial computer aided language learning (CALL) systems that employ speech tend to overestimate the reliability of the speech recognition technology [22]. To support learner speech recognition in the TLTS, our initial efforts focused on acoustic modeling for robust speech recognition especially in light of limited domain data availability [19]. In this case, we bootstrapped data from English and modern standard Arabic and adapted it to Levantine Arabic speech and lexicon. Dynamic switching of recognition grammars was also implemented, as were recognition confidence estimates, used by the pedagogical agent to decide how to give feedback. The structures of the recognition networks are distinct for the MSB and the MPE environments. In the MSB mode, the recognition is based on limited vocabulary networks with pronunciation variants and hypothesis rejection. In the MPE mode, the recognizer supports less constrained user inputs, focusing on recognizing the learner's intended meaning.

4.3 Mission Skill Builder Architecture

The Mission Skill Builder (MSB) is a one-on-one tutoring environment which helps the learner to acquire mission-oriented vocabulary, pronunciation training and gesture recognition knowledge. In this learning environment the learner develops the necessary skills to accomplish specific missions. A virtual tutor provides personalized feedback to improve and accelerate the learning process. In addition, a progress report generator generates a summary of skills the learner has mastered, which is presented to the learner in the same environment.

The Mission Skill Builder user interface is implemented in SumTotal's ToolBook, augmented by the pedagogical agent and speech recognizer. The learner initiates speech input by clicking on a microphone icon, which sends a "start" message to the automated speech recognition (ASR) process. Clicking the microphone icon again sends a "stop" message to the speech recognition process, which then analyzes the speech and sends the recognized utterance back to the MSB. The recognized utterance, together with the expected utterance, is passed to the Pedagogical Agent,

which in turn passes this information to the Error Model (part of the Language Model), to analyze and detect types of mistakes. The results of the error detection are then passed back to the Pedagogical Agent, which decides what kind of feedback to choose, depending on the error type and the learner's progress. The feedback is then passed to the MSB and is provided to the learner via the virtual tutor persona, realized as a set of video clips, sound clips, and still images. In addition the Mission Skill Builder informs the learner model about several learner activities with the user interface, which help to define and extend the individual learner profile.

4.4 Mission Practice Environment Architecture

The Mission Practice Environment (MPE) is responsible for realizing dramatically and visually engaging 3D simulations of social situations, in which the learner can interact with non-player characters by speaking and choosing gestures. Most of the MPE work is done in two modules: The Mission Engine and the Unreal World (see Figure 5). The former controls what happens while the latter renders it on the screen and provides a user interface.

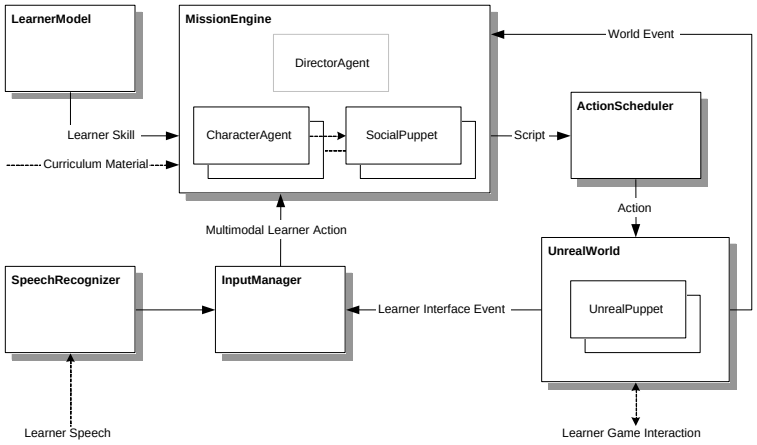

Fig. 5. The Mission Practice Environment architecture

The Unreal World uses the Unreal Tournament 2003 game engine where each character, including the learner's own avatar, is represented by an animated figure called an Unreal Puppet. The motion of the learner's puppet is for the most part driven by input from the mouse and keyboard, while the other puppets receive action requests from the Mission Engine through the Unreal World Server, which is an extended version of the Game Bots server [12]. In addition to relaying action requests to puppets, the Unreal World Server sends information about the state of the world back to the Mission Engine. Events from the user interface, such as mouse button presses, are first processed in the Input Manager, and then handed to the Mission Engine

where a proper reaction is generated. The Input Manager also invokes the Speech Recognizer, when the learner presses the right mouse button, and sends the recognized utterance, with information about the chosen gesture, to the Mission Engine.

The Mission Engine uses a multi-agent architecture where each character is represented as an agent with its own goals, relationships with other entities (including the learner), private beliefs and mental models of other entities [16]. This allows the user to engage in a number of interactions with one or more characters that each can have their own, evolving attitude towards the learner. Once the character agents have chosen an action, they pass their communicative intent to corresponding Social Puppets that plan a series of verbal and nonverbal behavior that appropriately carry out that intent in the virtual environment. We plan to incorporate a high-level Director Agent that influences the character agents, to control how the story unfolds and to ensure that pedagogical and dramatic goals are met. This agent exploits the learner model to know what the learners can do and to predict what they might do. The director will use this information as a means to control the direction of the story by manipulating events and non-player characters as needed, and to regulate the challenges presented to the student.

A special character that aides the learners during their missions uses an agent model of the learner to suggest what to say next when the learner asks for help or when the learner seems to be having trouble progressing. When such a hint is given, the Mission Engine consults the Learner Model to see whether the learner has mastered the skills involved in producing the phrase to be suggested. If the learner does not have the required skill set, the aide spells out in transliterated Arabic exactly what needs to be said, but if the learner should know the phrase in Arabic, the aide simply provides a hint in English such as "You should introduce yourself."

5 Evaluation

System and content evaluation is being conducted systematically, in stages. Usability and second language learning experts have evaluated and critiqued the learner interface, content organization, and instructional methods. Learner speech data are being collected, to inform and train the speech recognition models. Learners at the US Military Academy and at USC have worked through the first set of lessons and scenes and provided feedback. A formative evaluation study with eight beginning learners was performed in Spring 2004. Learners worked with the Tactical Language Training System in one 2-hour session. The subjects found the MPE game to be fun and interesting, and were generally confident that with practice, they would be able to master the game. This supports our hypothesis that the TLTS will enable a wide range of learners, including those with low levels of confidence, to acquire communication skills in difficult languages such as Arabic. However, the learners were generally reluctant to start playing the game, because they were afraid that they would not be able to communicate successfully with the non-player characters. To address this problem, we are modifying the content in the MSB to give learners more conversational practice and encourage learners to enter the MPE right away.

The evaluation also revealed problems in the MSB Tutoring Agent's interaction. The agent applied a high standard for pronunciation accuracy, which beginners found difficult to meet. At the same time, inaccuracies in the speech analysis algorithms caused the agent in some cases to reject utterances that were pronounced correctly. The algorithm for scoring learner pronunciation has since been modified, to give higher scores to utterances that are pronounced correctly but slowly; this eliminated most of the problems of correct speech being rejected. We have also adjusted the feedback selection algorithm to avoid criticizing the learner when speech recognition confidence is low. This revised feedback mechanism is scheduled to be evaluated in further tests with soldiers in July 2004 at Ft. Bragg, North Carolina.

6 Conclusions and Future Work

The Tactical Language Training System project has been active for a relatively brief period, yet it has already made rapid progress in combining pedagogical agent, pedagogical drama, speech recognition, and game technologies in support of language learning. Once the system design is updated based upon the results of the formative evaluations, the project plans the following tasks:

- integrate the Medina authoring tool to facilitate content development,
- incorporate automated tracking of learner focus of attention, to detect learner difficulties and provide proactive help,
- construct additional content to cover a significant amount of spoken Arabic,
- perform summative evaluation of the effectiveness of the TLTS in promoting learning, and analysis of the contribution of TLTS component features to learning effectiveness, and
- support translingual authoring – adapting content from one language to another, in order to facilitate the creation of similar learning environments for a range of less commonly taught languages.

7 Acknowledgments

The project team includes, in addition to the authors, CARTE members Catherine M. LaBore, David V. Pynadath, Nicolaus Mote, Shumin Wu, Ulf Hermjakob, Mei Si, Nadim Daher, Gladys Saroyan, Hartmut Neven, Chirag Merchant and Brett Rutland. From the US Military Academy COL Stephen Larocca, John Morgan and Sherri Bellinger. From the USC School of Engineering Shrikanth Narayanan, Naveen Srinivasamurthy, Abhinav Sethy, Jorge Silva, Joe Tepperman and Larry Kite. From the USC School of Education Harold O'Neil and Sunhee Choi, and from UCLA CRESST Eva Baker. Thanks to Lin Pirolli for her editorial comments. This project is part of the DARWARS initiative sponsored by the US Defense Advanced Research Projects Agency (DARPA).

References

1. Bernstein, J., Najmi, A. & Ehsani, F.: Subarashii: Encounters in Japanese Spoken Language Education. CALICO Journal 16 (3) (1999) 361-384
2. Brown, P. & Levinson: *Politeness: Some universals in language use*. Cambridge University Press, New York (1987)
3. DeSmedt, W.H.: Herr Kommissar: An ICALL conversation simulator for intermediate German. In V.M. Holland, J.D. Kaplan, & M.R. Sams (Eds.), *Intelligent language tutors: Theory shaping technology*, 153-174. Lawrence Erlbaum, Mahwah, NJ (1995)
4. Doughty, C.J. & Long, M.H.: Optimal psycholinguistic environments for distance foreign language learning. Language Learning & Technology 7(3), (2003) 50-80
5. Gamper, G. & Knapp, J.: A review of CALL systems in foreign language instruction. In J.D. Moore et al. (Eds.), Artificial Intelligence in Education, 377-388. IOS Press, Amsterdam (2001)
6. Gee, P.: What video games have to teach us about learning and literacy. Palgrave Macmillan, New York (2003)
7. Hamberger, H.: Tutorial tools for language learning by two-medium dialogue. In V.M. Holland, J.D. Kaplan, & M.R. Sams (Eds.), Intelligent language tutors: Theory shaping technology, 183-199. Lawrence Erlbaum, Mahwah, NJ (1995)
8. Harless, W.G., Zier, M.A., and Duncan, R.C.: Virtual Dialogues with Native Speakers: The Evaluation of an Interactive Multimedia Method. CALICO Journal 16 (3) (1999) 313-337
9. Holland, V.M., Kaplan, J.D., & Sabol, M.A.: Preliminary Tests of Language Learning in a Speech-Interactive Graphics Microworld. CALICO Journal 16 (3) (1999) 339-359
10. Johnson, W.L.: Interaction tactics for socially intelligent pedagogical agents. IUI 2003, 251-253. ACM Press, New York (2003)
11. Johnson, W.L.,& Rizzo, P.: Politeness in tutoring dialogs: "Run the factory, that's what I'd do." ITS 2004, in press (2004)
12. Kaminka, G.A., Veloso, M.M., Schaffer, S., Sollitto, C., Adobbati, R., Marshall, A.N., Scholer, A. and Tejada, S.: GameBots: A Flexible Test Bed for Multiagent Team Research. Communications of the ACM, 45 (1) (2002) 43-45
13. LaRocca, S.A., Morgan, J.J., & Bellinger, S.: On the path to 2X learning: Exploring the possibilities of advanced speech recognition. CALICO Journal 16 (3) (1999) 295-310
14. Lightbown, P.J. & Spada, N.: How languages are learned. Oxford University Press, Oxford (1999)
15. Marsella, S., Johnson, W.L. and LaBore, C.M.: An interactive pedagogical drama for health interventions. In Hoppe, U. and Verdejo, F. eds., Artificial Intelligence in Education. IOS Press, Amsterdam (2003)
16. Marsella, S.C. & Pynadath, D.V.: Agent-based interaction of social interactions and influence. Proceedings of the Sixth International Conference on Cognitive Modelling, Pittsburgh, PA (2004)
17. Muskus, J.: Language study increases. Yale Daily News, Nov. 21, 2003
18. NCOLCTL: National Council of Less Commonly Taught Languages. http://www.councilnet.org (2003)
19. Srinivasamurthy, N. and Narayanan: ``Language-adaptive Persian speech recognition", Proc. Eurospeech (Geneva,Switzerland) (2003)
20. Swartout, W.,Gratch, J., Johnson, W.L., et al. : Towards the Holodeck: Integrating graphics, sound, character and story. Proceedings of the Intl. Conf. on Autonomous Agents,409-416. ACM Press, New York (2001)
21. Swartout, W. & van Lent: Making a game of system design. CACM 46(7) (2003) 32-39

22. Wachowicz, A. and Scott, B.: Software That Listens: It's Not a Question of Whether, It's a Question of How. CALICO Journal 16 (3), (1999) 253-276
23. Witt, S. & Young, S.: Computer-aided pronunciation teaching based on automatic speech recognition. In S. Jager, J.A. Nerbonne, & A.J. van Essen (Eds.), Language teaching and language technology, 25-35. Swets & Zeitlinger, Lisse (1998)