Chapter 29
Second Language Learners' Spoken Discourse:
Practice and Corrective Feedback through Automatic Speech Recognition

Catia Cucchiarini
Radboud University, The Netherlands

Helmer Strik
Radboud University, The Netherlands

ABSTRACT

This chapter examines the use of Automatic Speech Recognition (ASR) technology in the context of Computer Assisted Language Learning (CALL) and language learning and teaching research. A brief introduction to ASR is first provided, to make it clear why and how this technology can be used to the benefit of learning and development in second language (L2) spoken discourse. This is followed by an overview of the state of the art in research on ASR-based CALL. Subsequently, a number of relevant projects on ASR-based CALL conducted at the Centre for Language and Speech Technology of the Radboud University in Nijmegen (the Netherlands) are presented. Possible solutions and recommendations are discussed given the current state of the technology with an explanation of how such systems can be used to the benefit of Discourse Analysis research. The chapter concludes with a discussion of possible perspectives for future research and development.

INTRODUCTION

Research on L2 learning has indicated that although exposure to the target language and usage-based learning are essential elements in the learning process, these are not always sufficient to guarantee target-like proficiency (Ellis, 2008; Ellis & Bogart, 2007). Focus on linguistic form provided through corrective feedback may help improve form accuracy in L2 spoken discourse. Unfortunately, in traditional teacher-fronted lessons there is generally not enough time for sufficient practice and feedback on speaking performance.

DOI: 10.4018/978-1-4666-6042-7.ch029

In this setting, the interest in applying ASR technology to L2 learning has been growing considerably in recent years (Eskenazi, 2009). ASR-based CALL systems would make it possible to offer sufficient amounts of practice in L2 speaking and to provide automatic feedback on different aspects of L2 spoken discourse. In this sense, ASR-based CALL systems would constitute an interesting supplement to traditional L2 classes. In addition, such systems can provide speaking practice in a private environment, which is a considerable advantage as speaking tasks are known to cause anxiety in L2 learners (Young, 1990). Moreover, L2 learners can practice at their own pace whenever they want.

In light of these advantages, research and development in this field have increased in recent years, and many systems have been developed that provide different forms of feedback on a variety of aspects of L2 spoken discourse. The majority of systems with a speech interactive nature address L2 pronunciation, which is considered a particularly challenging skill in L2 learning. A comprehensive overview of ASR-based commercial systems for L2 pronunciation is provided by Witt (2012). Many of these systems, however, do not contain important and desirable features of feedback on L2 pronunciation, such as immediate, detailed feedback on individual segments in the context of meaningful communicative tasks involving connected speech. In addition, CALL systems that are intended for practicing grammar skills and for improving accuracy in general do not support spoken interaction, but tend to resort to drag-and-drop exercises and typing (Bodnar, Cucchiarini, & Strik, 2011).

Against this background a number of projects were started at our lab which were aimed at conducting research and developing technology that would be conducive to the realization of ASR-based CALL systems that support practice and automatic feedback on L2 spoken discourse in line with insights from L2 learning research and L2 learners' requirements.

The aim of this chapter is to inform the reader about recent developments in the field of ASR-based CALL research and to indicate how these can lead to new methods and paradigms for the acquisition of spoken discourse in a second language. We first provide a brief introduction to ASR, to make it clear for the reader why and how this technology can be used to the benefit of learning and development in L2 spoken discourse. We then go on to provide an overview of the state of the art in research on ASR-based CALL. Subsequently, we present a number of relevant projects conducted at our lab and discuss possible solutions and recommendations for the development of ASR-based CALL systems and for the use of such systems to the benefit of Discourse Analysis. We then conclude with a discussion of possible perspectives for future research and development.

BACKGROUND: AUTOMATIC SPEECH RECOGNITION (ASR)

Standard ASR systems are generally employed to recognize words. The ASR system consists of a decoder (the search algorithm) and three 'knowledge sources': the language model, the lexicon, and the acoustic models. The language model (LM) contains probabilities of words and sequences of words. Acoustic models are models of how the sounds of a language are pronounced; in most cases so-called hidden Markov models (HMMs) are used, but it is also possible to use artificial neural networks (ANNs). The lexicon is the connection between the language model and the acoustic models. It contains information on how the words are pronounced, in terms of sequences of speech sounds. Therefore, the lexicon contains two representations for every entry: an orthographic transcription representing how a word is written and a phonological transcription representing how a word is pronounced. Since words can be pronounced in different ways, lexicons often contain more than one entry for some words, i.e. the