Available online at www.sciencedirect.com
ScienceDirect
Computer Speech and Language 25 (2011) 128–139
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

Review

Some background on dialogue management and conversational speech for dialogue systems

Yorick Wilks [a], Roberta Catizone [b,*], Simon Worgan [b], Markku Turunen [c]

[a] *Oxford Internet Institute, University of Oxford, Oxford, UK*
[b] *Computer Science Department, University of Sheffield, 211 Portobello Street, Sheffield S1 4DP, UK*
[c] *University of Tampere, Tampere, Finland*

Received 8 September 2009; accepted 9 March 2010
Available online 20 May 2010

Keywords: Dialogue systems; Human–computer interaction; Dialogue management; Dialogue architectures; Emotion detection

Contents

* Corresponding author. Tel.: +44 7788182975.
E-mail addresses: yorick.wilks@oii.ac.uk (Y. Wilks), roberta@dcs.shef.ac.uk (R. Catizone), simon.worgan@gmail.com (S. Worgan), mturunen@cs.uta.fi (M. Turunen).

0885-2308/$ – see front matter
doi:10.1016/j.csl.2010.03.001

1. Introduction

This special issue of the Journal is concerned with speech and language processing issues in the overall environment of end-to-end dialogue systems, and in particular with the sorts of techniques deployed in the COMPANIONS project (www.companions-project.org) which most of the contributors to this issue are associated in one way or another. The aim of the COMPANIONS project was to produce multimodal dialogue agent demonstrators within four years, and the papers in this volume that originate in that project are, in effect, two year prototypes, submitted to evaluations but designed principally as platforms (separately or by a new fusion of components) for further research on the deployment of emotion modelling and of machine learning (ML) techniques of a variety of forms. As will be described, there is already some reportable ML activity in these two-year prototypes.

COMPANIONS was also a much broader concept, embracing both the notion of a new form of conversational interface to the internet, while drawing on some of the traditions of the Embodied Conversational Agent (ECA); this tradition (e.g. Nagao and Takeuchi, 1994; Traum and Rickel, 2002) has developed rich models going beyond the basis "talking head" of its early days, but is nevertheless not at its heart a form of HLT (Human Language Technology) research and development. It is on this latter strand that the papers in the volume concentrate, along with the assumption that much research on emotion and politeness is far more dependent on language than its originators realise, and that specifically language and speech phenomena may be the best place to locate emotion and politeness—both crucial to a Companion—as opposed to say facial expressions and gestures, which are at the core of ECA work.

This initial paper surveys work in two areas: first, dialogue management (DM) which is at the core of the language processing system and extends from the understanding of input, in symbolic transcribed form, to decisions based on reasoning as to what to say next, right up to decisions about how to reply. Here we shall concentrate mainly on the core DM itself and its associated knowledge representation and reasoning. Secondly, we shall look very broadly at the speech recognition aspect of conversational speech: this is a very large area and we can only lay out very broad categories of work.

Dialogue systems have been around since the 1960s, the best known are conversation programs such as Eliza (Weizenbaum, 1966) and Parry (Colby, 1973). The approaches we describe are categorised as follows: finite state/dialogue grammars, plan-based and collaborative; however, this division is not perfect, since any system can in the end be implemented as a finite state system, but the distinction corresponds to design approaches versus implementation approaches, since finite state models can be used to implement a variety of approaches independently of the design choice. Again, collaborative models may or may not be plan-based, so this distinction too, is less than firm.

2. Basic types of dialogue management systems

2.1. Dialogue grammars and frames

Dialogue grammars, are systems that identify and represent local or global surface patterns of dialogue or patterns of speech acts (Searle, 1969) and their responses. Dialogue grammars, which have a long history (Polany and Scha, 1984; Reichman, 1981; Sinclair and Coulthard, 1975), use prescriptive grammars for pattern sequences in dialogues. The first grammars described the structure of the complete dialogue, from beginning to end, whereas more recent approaches are based on the observation that there are a number of sequencing regularities in dialogues, which are called *adjacency pairs*. It has been proposed that a dialogue is a collection of such pairs (Jefferson, 1972), which describe facts such as that questions are generally followed by answers, proposals by acceptances, denials, etc. Digressions and repairs are dealt with by using embedded sequences.

Dialogue grammars are used to parse the structure of a dialogue, just as syntactic grammar rules are used to parse sentences. Phrase-structure grammar rules and various kinds of state machines have been used to implement dialogue grammars. For example the SUNDIAL system, uses a dialogue grammar to engage in dialogue about travel conversations.

Although dialogue grammars have been successfully implemented (Müller and Runger, 1993; Nielsen and Baekgaard, 1992), they have been criticised on the grounds that they lack flexibility both as to deviations in the dialogue as well as portability to other domains.

A significant extension of dialogue grammars are *frame-based approaches*, which have been developed to overcome the lack of flexibility of dialogue grammars. The entities in the application domain are hierarchically modelled, and the