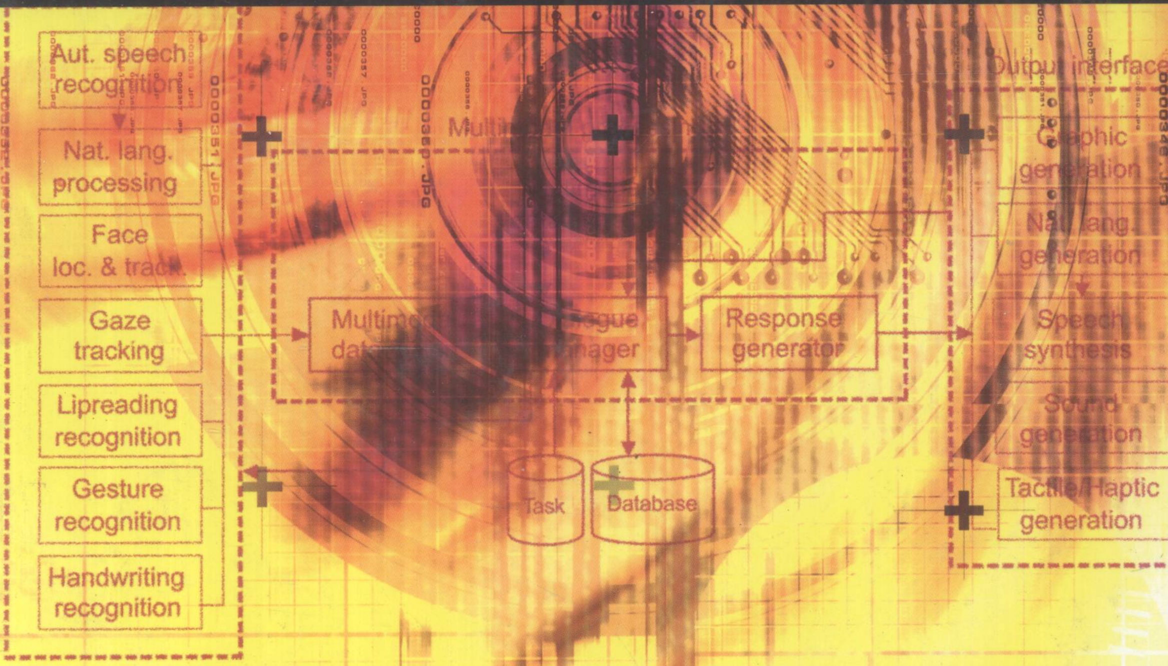
SPOKEN, MULTILINGUAL AND MULTIMODAL DIALOGUE SYSTEMS

DEVELOPMENT AND ASSESSMENT



RAMÓN LÓPEZ-CÓZAR DELGADO

MASAHIRO ARAKI

WILEY

SPOKEN, MULTILINGUAL AND MULTIMODAL DIALOGUE SYSTEMS

DEVELOPMENT AND ASSESSMENT

Ramón López-Cózar Delgado

Granada University
Spain

Masahiro Araki

Kyoto Institute of Technology
Japan

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears
in print may not be available in electronic books.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-470-02155-2
ISBN-10 0-470-02155-1

Typeset in 10/12pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

SPOKEN, MULTILINGUAL AND MULTIMODAL DIALOGUE SYSTEMS

Preface

In many situations, the dialogue between two human beings seems to be performed almost effortlessly. However, building a computer program that can converse in such a natural way with a person, on any task and under any environmental conditions, is still a challenge. One reason why is that a large amount of different types of knowledge is involved in human-to-human dialogues, such as phonetic, linguistic, behavioural, cultural codes, as well as concerning the world in which the dialogue partners live. Another reason is the current limitations of the technologies employed to obtain information from the user during the dialogue (speech recognition, face localisation, gaze tracking, lip-reading recognition, handwriting recognition, etc.), most of which are very sensitive to factors such as acoustic noise, vocabulary, accent, lighting conditions, viewpoint, body movement or facial expressions. Therefore, a key challenge is how to set up these systems so that they are as robust as possible against these factors.

Several books have already appeared, concerned with some of the topics addressed in this book, primarily speech processing, since it is the basis for spoken dialogue systems. A huge amount of research papers can be found in the literature on this technology. In recent years, some books have been published on multimodal dialogue systems, some of them as a result of the selection of workshops and conference papers. However, as far as we know, no books have as yet been published providing a coherent and unified treatment of the technologies used to set up spoken, multilingual and multimodal dialogue systems. Therefore, our aim has been to put together all these technologies in a book that is of interest to the academic, research and development communities.

A great effort has been made to condense the basis and current state-of-the-art of the technologies involved, as well as their technological evolution in the past decade. However, due to obvious space limitations, we are aware that some important topics may have not been addressed, and that others may have been addressed quite superficially. We have tried to speak to all the constituencies. The topics cover a wide range of multi-disciplinary issues and draw on several fields of study without requiring too deep an understanding of any area in particular; in fact, the number of mathematical formulae is kept to a minimum. Thus, we think reading this book can be an excellent first step towards more advanced studies.

The book will also be useful for researchers and academics interested in having some reference material showing the current state-of-the-art of dialogue system. It can also be useful for system developers interested in exploiting the emerging technology to develop automated services for commercial applications. In fact, it contains a large number of Internet links where the reader can find detailed information, development Internet sites and development tools download. Professors as well as undergraduate and post-graduate students of Computer Science, Linguistics, Speech and Natural Language Processing, Human-Computer Interaction and Multimodal Interactive Systems will also find this text useful.

Writing this book has been a challenging and fascinating journey down many endless roads, full of huge amounts of information concerned with the technologies addressed. As commented above, given the limitations of space, it has not been easy to come up with a trade-off between discussing the diverse topics in detail, on the one hand, while, on the other, giving a general, wide-range overview of the technologies. We hope the efforts we made to condense this universe of information into just one book will benefit the reader.

Journeying down these roads we have encountered many researchers and companies that have kindly granted permissions to reproduce material in this book. We wish to thank the kind collaboration of them all, and especially the contribution of Jan Alexandersson (DFKI, Germany), Koray Balci (ITC-irst Cognitive and Communication Technologies Division, Italy), Marc Cavazza (University of Teesside, UK), Mark Core (University of Southern California, USA), Jens Edlund (KTH, Sweden), James R. Glass (MIT, USA), Clegg Ivey (Voxeo Corporation, USA), Sanshzar Kettebekov (Advanced Interfaces, Inc., USA), Michael F. McTear (University of Ulster, Northern Ireland), Nick Metianu (IBM Software Group, USA), Yasuhisa Niimi (ATR, Japan), Rainer Stiefelhagen (University of Karlsruhe, TH, Interactive Systems Labs, Germany), Kevin Stone (BeVocal Café, USA), Jan Van Santen (OGI School of Science and Engineering, Oregon Health and Science University (USA), and Yunbiao Xu (Hangzhou University of Commerce, China).

We also would like to thank the AAAI, Elsevier and Springer for their kind permission to reproduce material in this book.

Finally, we would also like to thank very much and in particular the support, help and contribution of the student and scholarship holder, Zoraida Callejas, and Professor Miguel Gea in the Department of Languages and Computer Science at Granada University, Spain.

Ramón López-Cózar Delgado
Granada

Masahiro Araki
Kyoto
April 2005

Contents

1

Introduction to Dialogue Systems

1.1 Human-Computer Interaction and Speech Processing

The so-called Human-Computer Interaction (HCI) is a multidisciplinary field in which three main elements are involved: human beings, computers and interaction. Research on HCI is very important because it stimulates the development of new interfaces that reduce the complexity of the interaction and ease the use of computers by non-expert users. In the past, HCI was extremely rigid since the user had to interpret the information provided by the computer expressed in a very different language from the human one. Technological advances have greatly improved the interaction. For example, the first-generation computers that only allowed letters to be displayed have been replaced by multimedia computers that allow reproduction of graphics, videos, sounds, etc., making the interaction much more comfortable. However, classical interaction with computers based on screen, keyboard and mouse, can be carried out only after the user has the minimal knowledge about hardware and software.

An alternative and relatively new way of interacting with computers is based on the processing of the human speech, which allows several advantages in comparison to the classical one based on keyboard, mouse and screen. Among others, speech offers a greater speed for transmitting information, allows other tasks (liberating the user from the need to use his or her hands and/or eyes) to be carried out simultaneously, reveals the identity of the speaker and permits some disabled users to interact with the computer. Speech allows a great expressivity, in fact, human beings express their ideas, feelings, etc. in language. Speech also allows information about the state of mind of the speaker, his/her attitude towards the listener, etc. to be transmitted. Moreover, speech can be transmitted by simple and widely used devices such as fixed and mobile telephones, making possible remote access to a variety of speech-based services.

The start of speech-based interaction with computers can be traced back to 1977, when in the USA several companies started to develop commercial applications at a very low cost, as, for example, the speaking calculator presented by Telesensory Systems Inc. or the Speak'n Spell system by Texas Instruments. Among other applications, this kind of interaction is currently used to interact with program interfaces (e.g. to move the cursor to a

Spoken, Multilingual and Multimodal Dialogue Systems: Development and Assessment Ramón López-Cózar Delgado and Masahiro Araki © 2005 John Wiley & Sons, Ltd

specific position of the screen) and operating systems (e.g. to run programs, open windows, etc.). This type of interaction is also used in dictation systems, allowing the user to write documents without the need to type but only say the words. Speech-based communication is also used to control devices in domestic environments (e.g. turn on lights, ovens, hi-fi sets, etc.) which can enhance the quality of life of disabled people. Moreover, this kind of communication is used to interact with car navigation and other in-car devices, allowing a hands- and eye-free interaction for drivers that increases their safety.

1.2 Spoken Dialogue Systems

Another kind of application of the speech-based interaction is the so-called Spoken Dialogue Systems (SDSs), also called *conversational systems*, that can be defined as computer programs developed to provide specific services to human beings in the same way as if these services were provided by human beings, offering an interaction as natural and comfortable as possible, in which the user interacts using speech. It could be said that the main feature of these systems is their aim to behave 'intelligently' as if they were human operators in order to increase the speed, effectiveness and ease of obtaining specific services automatically. For that purpose, these systems typically include a module that implements the 'intelligence' of the human being whose behaviour they aim to replace in order to provide users with a natural and effective interaction. Among other applications, these systems have been used to provide automatic telephone services such as airplane travel information (Seneff and Polifroni 2000), train travel information (Billi et al. 1997; Torres et al. 2003; Vilar et al. 2003), weather forecasts (Zue et al. 2000; Nakano et al. 2001), fast food ordering (Seto et al. 1994; López-Cózar et al. 1997), call routing (Riccardi et al. 1997; Lee et al. 2000), and directory asistance (Kellner et al. 1997).

The use of dialogue systems has increased notably in recent years, mainly due to the important advances made by the Automatic Speech Recognition (ASR) and speech synthesis technologies. These advances have allowed the setting up of systems that provide important economic savings for companies, offering an automatic service available 24 hours a day to their customers. The initial systems were very limited with regard to the types of sentences that could be handled and the types of task performed but in the past three decades the communication allowed by this kind of system has improved notably in terms of *naturality* or similarity to human-to-human communication. However, the dialogue between human beings relies on a great diversity of knowledge that allows them to make assumptions and simplify the language used. This makes it very difficult for current dialogue systems to communicate with human beings in the same way humans carry on a dialogue with each other.

Although the functionality of these systems is still limited, some of them allow conversations that are very similar to those carried out by human beings, they support natural language phenomena such as anaphora and ellipsis, and are more or less robust against spontaneous speech phenomena as, for example, lack of fluency, false starts, turn overlapping, corrections, etc. To achieve robustness in real-world conditions and portability between tasks with little effort, there is a trend towards using simple dialogue models when setting up these systems (e.g. state transition networks or dialogue grammars) and using simple representations of the domains or tasks to be carried out. This way, it is possible to use information regarding the likely words, sentence types and user intentions. However, there are also proposals on using much more complex approaches. For example, there are models based on Artificial

Intelligence principles that emphasise the relationships between the user's sentences and his plans when interacting with a human operator or an automatic system, and the importance of applying reason to the beliefs and intentions of the dialogue partners. There are also hybrid models between both approaches that attempt to use the simple models enhanced with specific knowledge about the application domain, or that include plan inference strategies restricted to the specific application of the system.

Speech is the most natural communication means between human beings and is the most adequate communication modality if the application requires the user to have his eyes and hands occupied carrying out other tasks, as, for example, using a mouse, a keyboard, or driving a car. However, dialogue systems based exclusively on speech processing have some drawbacks that can result in less effective interactions. One is derived from the current limitations of the ASR technology (Rabiner and Juang 1993), given that even in very restricted domains and using small vocabularies, speech recognisers sometimes make mistakes. In addition to these errors, users can utter out-of-domain words or sentences, or words of the application domain not included in the system vocabulary, which typically causes speech recognition errors. Hence, to prevent these errors in the posterior analysis stages, the systems must confirm the data obtained from the users. Another problem, specially observed in telephone-based dialogue systems that provide train or airplane information, is that some users may have problems understanding and taking note of the messages provided by the systems, especially if the routes take several transfers or train connections (Claasen 2000). Finally, it has also been observed that some users may have problems understanding the possibilities of the system and the dialogue status, which causes problems of not knowing what to do or say.

1.2.1 Technological Precedents

SDSs offer diverse advantages in comparison to previous technologies developed to interact with remote, interactive applications that provide information or specific services to users. For example, one of these technologies is Dual Tone Multiple Frequency (DTMF), in which the user interacts with the system by pressing keys on the telephone that represent functions of the application (e.g. 1 = accept, 2 = deny, 3 = finish, etc.). Although this type of interaction may be appropriate for applications with a small number of functions, dialogue systems allow much more flexibility and expression power using speech, as users do not need to remember the assignation of keys to the functions of the application.

Another way to communicate with remote applications, before the advent of the current dialogue systems, is based on using speech in the form of isolated words. This way it is possible to use the words 'yes' or 'no', for instance, instead of pressing keys on the telephone. This technology has been implemented in applications with small vocabularies and a small number of nested functions. In comparison to this technology, dialogue systems offer the same advantages mentioned before, given that in these applications speech does not allow any expression facility and is a used merely as an alternative to pressing telephone keys.

Another technology developed to interact with remote applications using speech is based on the so-called *sub-languages*, which are subsets of natural language built using simple grammars (Grishman and Kittredge 1986). Following this approach, Sidner and Forlines (2002) developed a sub-language using a grammar that contains fourteen context-free rules to parse very simple sentences, mainly formed by a verb in the imperative form followed by a noun or a pronoun. This sub-language was used to interact with the Diamond Talk system developed to record and playback television programs. The system receives the user voice, processes it and generates an output both in speech form (using a speech synthesiser) and text format (on the TV screen).

Compared with this technology, dialogue systems provide a more natural interaction since users are not forced to learn specific languages: they can use their own natural language. Thus, these systems facilitate the query of databases, specially for non-expert users who can carry out queries expressed in natural language instead of having to learn database-specific languages as SQL. Also, these systems allow the users to refine and modify the queries by means of a continuous dialogue with the system, instead of having to build the queries in just one command.

1.3 Multimodal Dialogue Systems

Dialogue systems can be classified according to several criteria. One is the communication modalities (or channels) used during the interaction. Regarding this criterion, these systems can be classified into *unimodal* and *multimodal*. In a unimodal system the information is transmitted using just one communication channel; this is the case to SDSs since the interaction is carried out using only speech. These systems have several drawbacks mainly derived from the current limitations of Automatic Speech Recognition (ASR) technology. In order to solve these limitations, in the past few years there has been a notable research interest in the joint utilisation of speech and other communication modalities, leading to the so-called Multimodal Dialogue Systems (MMDSs) that are studied in Chapter 3.

Multimodality refers to the use of more than just one communication channel to transmit/receive information to/from the user (e.g. speech, gestures, gaze, facial expressions, etc.), which allows a reduction in the number of interaction errors. In fact, human-to-human communication relies on several communication modalities such as, for example, speech, gaze, body gestures, facial expressions, etc. Human beings use all these information sources (unconsciously on many occasions) to add, modify or substitute information in the speech-based communication, which allows effective recognition and understanding rates even in noisy environments. MMDSs aim to replicate the human-to-human communication which is in essence a multimodal process, given that several communications channels take part. Thus, multimodal systems are designed to support more transparent, flexible, effective and powerful interaction in a wider range of applications, to be used by non-expert users and to offer a more robust performance in changing environmental conditions. Among other applications, these systems have been developed to provide information about microwave ovens (Beskow and McGlashan 1997), available apartments (Gustafson et al. 1999, 2000; Bickmore and Cassell 2004), or boat traffic conditions (Beskow 1997), and have also been applied to the interaction with mobile robots (Lemon et al. 2001), and appointment scheduling (Vo and Wood 1996).

The so-called *multimedia* systems also use several communication channels to transmit/receive information. For example, they can transmit sounds and speech to the user through the loudspeakers, and present graphic and written information on the screen. Also, they can receive information from the user through the keyboard, or a microphone, etc. However, the difference between both types of system is that the multimodal one processes information at a higher abstraction level in order to obtain its meaning (i.e. the user intention) whereas this high-level information representation is unnecessary for multimedia systems.

In this book we consider that any dialogue system can be seen as a particular case of MMDS that uses more or less interaction modalities, depending on the type of dialogue system. Typically, the design of such a multimodal system considers the independence of dialogue management strategies, the internal representation of data provided by the user, the task to be carried out by the system, and the interaction modalities. This way a user may provide data to the system using a microphone, computer display, data glove or keyboard, for example, while the system may provide responses using any output device, as for example a loudspeaker or a display. The input provided by the user through any of the input devices is converted into an internal representation independent of the device. Thus, it makes no difference for the system where the input comes from, i.e. the user may either use a pen to click on a 'Cancel' button shown on the screen, utter a particular word associated with this action, or press the ESC key in the keyboard. This independence leads to a system architecture consisting of several modules that can be organised into three components: input interface, multimodal processing and output interface, together with the task and database modules (Figure 1.1). This figure shows the conceptual module structure we are using for explanation purposes throughout this book. The system modules are discussed with some detail in Chapter 2.

In the input interface (studied in Section 2.1) the ASR module analyses the user voice and transforms it into a sequence of words in text format. The natural language processing (NLP) module[1] analyses the sequence of words in order to obtain its meaning. The face localisation and tracking module finds the face of the user as soon as he or she enters into the vision range of a camera connected to the system, and tracks its movement in a sequence of images. The gaze tracking module processes the images provided by the camera to detect and track the movement of the user's eyes. The lip-reading recognition module detects and analyses the user's lip movements, which are very important to enhance the ASR specially when the speech signal is degraded by noise or cross-talk (i.e. other people talking near the microphone or the telephone). The gesture recognition module analyses the gestures

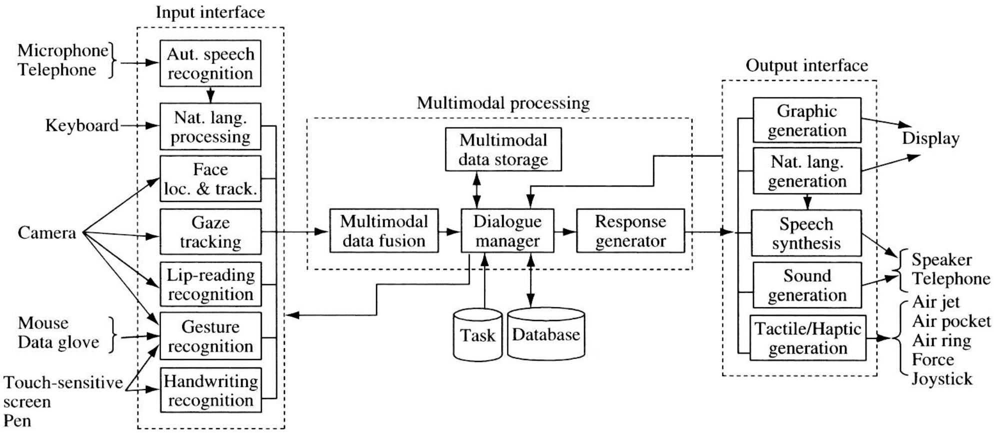


Figure 1.1 Conceptual structure of a multimodal dialogue system

[1] Some researchers call this module 'NLU' instead of 'NLP' as it is concerned with the Natural Language Understanding. Thus, in this book we call it either NLP or NLU.

made by the user to detect specific movements of some parts of the body (e.g. hands, arms, head, etc.) and to recognise determined communicative acts (e.g. affirm, negate, etc.). The handwriting recognition module deals with the input the user may provide by writing (using a stylus on a Personal Digital Assistant (PDA) screen, for instance). This input modality is specially important for applications that use foreign words (e.g. city names), which may be easier recognised written than spoken. In addition, this interaction modality can be used to provide additional vocabulary to the system.

In the multimodal processing component (studied in Section 2.2), the multimodal data fusion module receives the information chunks provided by the diverse input modules and combines them in order to obtain the intention of the user (operation called *fusion*). This module sends the obtained intention representation to the next module of the architecture, called the dialogue manager, that decides the next action of the system in order to carry out a particular task defined in the task module of the system. By changing the contents of this module the system can be ported to a different application domain, i.e. carry out a different task. Considering the task and the dialogue status, the dialogue manager can initiate a diversity of actions, for example, a query to the database module (either local or remote) that contains the information requested by the user, a prompt for additional data to query the database, a confirmation of data obtained from the user with little confidence, etc. The representation of the user intention is stored in the so-called multimodal data storage module, which stores all the interactions made by the user (and possibly the system) and provides the dialogue history for the dialogue manager to resolve possible ambiguities and referring expressions. Finally, the response generator module carries out the *fission* operation (as opposed to the fusion operation commented before), consisting in choosing the output to be produced in each output modality and coordinating the output across the modalities.

In the output interface (studied in Section 2.3), the graphic generation module presents information visually on the computer screen (e.g. graphics, pictures, icons, videos, etc.). It usually also includes a human-like character called the *animated agent* (studied in Section 3.2.4), that moves the lips and makes facial expressions and/or body gestures in synchronisation with the speech output. The natural language generation (NLG) module creates messages in text mode to be displayed on the screen and also text messages to be converted into the system voice through the speech synthesis module, which makes the text-to-speech conversion. The sound generation module generates specific sounds that can be useful in providing a more informative, friendly and comfortable interaction for the user, as, for example, a distinct sound if the sentence uttered cannot be understood by the system. Finally, the tactile/haptic module generates output information to stimulate the somatic senses of the user using tactile, haptic or force output devices.

It should be noted that Figure 1.1 presents a conceptual, general architecture whose modules have assigned the functions mentioned above. In some implemented systems some of these modules are subdivided into other modules, whereas in other systems, modules separated in this figure are integrated in just one module. For example, in the SmartKom system (Pfleger et al. 2003), the dialogue manager is divided into two modules: *discourse modeller* and *action planner*. The task of the former is to enrich the hypotheses generated by the data fusion module with information about the current context, validate these hypotheses, and collaborate with the fusion module in resolving referring expressions, whereas the task of the action planner is to access connected services and/or devices. However, in the Olga system (Beskow and McGlashan 1997), the dialogue manager not only decides the next action of the system but also decides and coordinates the response generation; thus, in this case the response generation module can be considered integrated into the dialogue manager.

In Figure 1.1, the arrow from the dialogue manager module to the input interface indicates that this module provides information that can be used by several modules of the interface. This information is typically concerned with the dialogue status and expectations about the input the user will likely make in his next turn, which may be very useful to enhance the performance of several recognition modules. For example, the expectations may be useful in selecting a particular language model used by the speech recognition module. Analogously, the arrow from the output interface to the dialogue manager indicates that the output of several modules in the interface can be used as feedback to the dialogue manager. For example, this information may be useful to resolve user references to objects previously shown on the system screen (studied in Section 5.4.5).

It can also be noted that considering the system architecture shown in Figure 1.1, an SDS can be considered as a particular case of an MMDS in which only the modules concerned with speech processing (ASR, NLP, NLG and speech synthesis), in addition to the dialogue manager, are used. In this case, the input is unimodal (only speech), the multimodal data fusion only considers data obtained from the uttered sentences, and the generation of responses is also unimodal (only speech and maybe other sounds).

1.4 Multilingual Dialogue Systems

Another criterion to classify dialogue systems is in terms of the number of languages supported. Considering this criterion, they can be classified into *monolingual* and *multilingual*. The latter solves a limitation of the former, namely that the interaction must be carried out using a specific language, which makes them unusable for users who do not speak that language. Multilingual Dialogue Systems (MLDSs), studied in Chapter 4, present the advantage of allowing the interaction using several languages (e.g. English, French, German, etc.). Among other applications, these systems have been applied to provide information about geographical areas, such as points of interest, distance between places, etc. (Glass et al. 1995), and weather information (Zue et al. 2000). In particular, these systems are very useful for some types of user, e.g. tourists who cannot speak the local language of the country they are visiting. Generally these systems are designed to be as language-independent as possible by storing language-dependent information in external configuration files. This development is based on the fact that it is possible to obtain the same semantic representation from the analysis of a given sentence expressed in different languages, in which only the language-dependent words change. This method of analysing sentences is similar to the *interlingua* approach to Machine Translation (MT) (Roe et al. 1991; Waibel et al. 1991; Hutchins and Somers 1992; Wahlster 1993, 1994). Given that the semantic representations do not change, the dialogue management strategies can be language-independent. Also, instead of using just one set of patterns to generate the sentences, these systems generally use several sets (one per language) which allow to generate the same kind of spoken sentences in different languages using a multilingual NLG module (for generating the sentences in text mode) and a multilingual speech synthesiser (to carry out the text-to-speech conversion).

1.5 Dialogue Systems Referenced in This Book

A large amount of dialogue systems (either spoken, multilingual or multimodal) for a great number of application domains can be found in the literature. Table 1.1 shows some of them, which have been mostly referenced throughout this book for explanation purposes.

Table 1.1 Dialogue systems mostly referenced in this book

System/Year/ Reference	Application domain	Lab	Input	Output	Languages	Remarks
AdApt Gustafson et al. 2000, 2002; Bell et al. 2000; Skantze 2002	Real estate	KTH and Telia Research (Sweden)	Speech, mouse, gestures	Speech, graphics, animated agent	Swedish	
August 1999 Gustafson et al. 1999	Info. about city facilities, KTH and life and works of A. Strindberg	KTH (Sweden)	Speech	Speech, text, animated agent	Swedish	Push-and-talk for ASR
Jeanie 1996 Vo and Wood 1996	Appointment scheduling	Carnegie Mellon Univ. (USA)	Speech, gestures, handwriting	Speech	English	
Jupiter 2000 Zue et al. 2000	Weather information	Lab for Computer Science MIT (USA)	Speech	Speech	English, German, Japanese, Mandarin Chinese, Spanish	
KIT system 2002 Xu et al. 2002	Sightseeing, hotel search, PC assembly	KIT (Japan)	Speech	Speech	Japanese, Chinese	
Kyoto voice portal 2005 Omukai and Araki 2004	Weather, bus, restaurant, sightseeing	KIT (Japan)	Speech	Speech, graphics	Japanese, English	