

# Appendixes

---

# Appendix 1:

## Grammar Formats

---

There are a number of different grammar formats that are currently supported on various VoiceXML platforms. The VoiceXML 1.0 specification used the Java Speech Grammar Format (JSGF) in its examples when illustrating the use of grammars, and JSGF is the format used in the current version of the IBM WebSphere Voice SDK (Speech Development Kit). The IBM Voice Toolkit also supports the Speech Recognition Grammar Format (SRG XML) and the Speech Recognition Control Language (SRCL), a variant of the Backus–Naur Form (BNF). Some VoiceXML platforms use proprietary grammar formats. For example, Nuance, Voxpilot and BeVocal support the Grammar Specification Language (GSL) format, which was developed by Nuance and is used widely in several VoiceXML textbooks.

In an effort to arrive at a standard for VoiceXML grammars, the VoiceXML 2.0 specification defines the XML form of the World Wide Web Consortium (W3C) Speech Recognition Grammar (SRG) Format as required for all VoiceXML 2.0 browsers. An alternative format, the Augmented BNF (ABNF) form, is defined as optional. Platforms can support additional formats if they so wish.

The XML form represents a grammar as an XML document with the logical structure of the grammar captured by XML elements. This format is ideal for computer-to-computer communication of grammars, as it is compatible with widely available XML technology, such as parsers and XSLT (Extensible Style Language Transformation).

The ABNF form combines traditional BNF Spoken Dialogue Technology with a regular expression language. This format is familiar to many speech application developers and is a more compact and more human-readable representation than XML. It is also fairly similar to the JSGF. Some platforms allow grammars written in one format to be converted to another format.

### The XML and ABNF Grammar Formats

There is a direct mapping between the XML and ABNF grammar formats, so that it is easy to convert a grammar written in one of these formats into the

**Table A1.1.** XML and ABNF grammar formats

ABNF grammar format	XML grammar format
Expansion involving further rules: \$viewdetails = \$studentname \$coursename	<pre>&lt;rule id = "viewdetails"&gt;   &lt;ruleref uri = "#studentname"&gt;   &lt;ruleref uri = "#coursename"&gt; &lt;/rule&gt;</pre>
Expansion involving choice of tokens: \$studentname = john   david;	<pre>&lt;rule id = "studentname"&gt;   &lt;one-of&gt;   &lt;item&gt; john &lt;/item&gt;   &lt;item&gt; david &lt;/item&gt; &lt;/one-of&gt; &lt;/rule&gt;</pre>
Expansion with optional and repeated items: Optional: [student]	<pre>&lt;count number = "optional"&gt;   student &lt;/count&gt;</pre>
Zero or more times: coursename*	<pre>&lt;count number = "0+ "&gt;   coursename &lt;/count&gt;</pre>
One or more times: coursename+	<pre>&lt;count number = "1+ "&gt;   coursename &lt;/count&gt;</pre>

**Table A1.2.** Semantic tags in XML and ABNF grammars

ABNF grammar format	XML grammar format
\$coursename = comms   communications {"01"}	<pre>&lt;rule id = "coursename"&gt;   &lt;one-of&gt;   &lt;item&gt; comms &lt;/item&gt;   &lt;item&gt; communications   &lt;tag&gt; "01" &lt;/tag&gt;   &lt;/item&gt;   &lt;/one-fo&gt; &lt;/rule&gt;</pre>

other format. Table 1 illustrates how different types of rule are expressed in the two formats.

## Semantic Tags

Semantic tags can be associated with rules, as described in Chapter 10. Table 2 illustrates semantic tags in the ABNF and XML grammar formats to cater for alternative phrasings. In this example, the coursename can be either “comms” or “communications”, and the value to be returned if either word is used is “01”.

## Headers and Grammar References

Headers that are grammar references and extensions vary slightly from one platform to another.

### IBM WebSphere

The XML format is supported on the IBM WebSphere platforms, but requires a header such as the following (note that the root rule has to be specified):

```
<grammar type="application/srgs+xml" root="choice">
```

The following is an inline XML grammar that could be used with the file “studentsystem1a” (see Chapter 9):

```
<grammar type="application/srgs+xml" root="choice">
<rule id = "choice" scope = "public">
<one-of>
<item> students </item>
<item> courses </item>
<item> reports </item>
</one-of>
</rule>
</grammar>
```

There are certain limitations on XML grammars in WebSphere Voice Server Version 3.1. For example, external references to grammars by the Uniform Resource Indicator (URI) and tag-format declarations are not supported. For this reason it is currently not possible to write form-level grammars in XML format for mixed-initiative forms for examples such as “studentsystem6a.vxml” (from Chapter 10) that will run on the WebSphere Voice Server platform.

### Bevocal

Bevocal supports the XML grammar format. The required header is:

```
<grammar type="application/grammar+xml" version = "1.0" root="choice">
```

Note that the version number is required.

Form-level grammars are supported for mixed-initiative forms. We can illustrate with the example “studentsystem6a” from Chapter 10, in which the user could state a `studentname` and a `coursename` within the same utterance. The following is the XML version of the grammar (“studentdetails.gram”). Note that the extension “.grxml” is required:

```
<grammar type="application/grammar+xml" xml:lang = "en" root="studentsystem6a" version="1.0">
<rule id="studentsystem6a" scope="public">
<ruleref uri="#studentname"/>
<ruleref uri="#coursename"/>
</rule>
<rule id="studentname" scope="public">
<one-of>
```

```

<item> john <tag> studentname="john" </tag> </item>
<item> david <tag> studentname="david" </tag> </item>
<item> rosemary <tag> studentname="rosemary" </tag> </item>
<item> jennifer <tag> studentname="jennifer" </tag> </item>
</one-of>
</rule>

<rule id="coursename" scope="public">
<one-of>
<item> communications <tag> coursename="communications" </tag> </item>
<item> algorithms <tag> coursename="algorithms" </tag> </item>
<item> programming <tag> coursename="programming" </tag> </item>
<item> databases <tag> coursename="databases" </tag> </item>
</one-of>
</rule>
</grammar>

```

This grammar is referenced in the document “studentstem6a.vxml” as follows:

Form-level grammar (calling the top-level rule “studentsystem6a”):

```

<grammar type="application/grammar+xml" version="1.0"
src="studentsystem6a.grxml#studentsystem6a"/>

```

Field-level grammar for “studentname”:

```

<grammar type="application/grammar+xml"
src="studentsystem6a.grxml#studentname"/>

```

Field-level grammar for “coursename”:

```

<grammar type="application/grammar+xml"
src="studentsystem6a.grxml#coursename"/>

```

Note that JSGF grammars can be used on the BeVocal platform, although they need to be enclosed within a CDATA section, as in the following example:

```

<grammar>
  <![CDATA[
    #JSGF 1.0;
    grammar topping;
    public <topping> =
      ( pepperoni | sausage | mushrooms );
  ]]>
</grammar>

```

Vxml documents also require a reference in the document header to the document type definition (DTD), as in the following example:

```

<?xml version="1.0"?>
<!DOCTYPE vxml PUBLIC "-//BeVocal Inc//VoiceXML 1.0//EN"
"http://cafe.bevocal.com/libraries/dtd/vxml1-0-bevocal.dtd">
<vxml version="1.0">

```

# Appendix 2:

## The CSLU Toolkit and Related Software for Chapters 7 and 8

---

In order to run and develop the exercises in Chapters 7 and 8 the following software is required:

1. CSLU toolkit.
2. TcLODBC.

You will also need to set up data sources using the ODBC Data Source Administrator.

### CSLU Toolkit

The CSLU toolkit is available from the following site:

<http://www.cslu.ogi.edu/toolkit/>

Full installation instructions are provided at the site. The toolkit currently runs under Windows 95, 98, NT, ME, 2000, XP Home and XP Professional on Intel (or compatible) processors only.

### TcLODBC

TcLODBC is required to make a connection between the CSLU toolkit and an ODBC compatible database, such as Microsoft Access, as used in Chapters 7–11. Information on downloading and installing TcLODBC can be found at: <http://sourceforge.net/projects/tclodbc>

Download the Windows version (tclodbc.2.3.1.zip) to a temporary directory. Click on tclodbc.2.3.1.zip and then click on the file “setup.tcl”. This will cause tclodbc to be installed in the directory Program Files\CSLU.

### Setting Up a Data Source (ODBC)

ODBC is used to access data from a variety of database management systems. To set up a data source in Windows XP:

Click **Start**, click **Control Panel**, and then click **Performance and Maintenance**. Click **Administrative Tools**, and then double-click **Data Sources (ODBC)**. This will bring up the ODBC Data Source Administrator. For further instructions, click on the **Help** tab in the bottom right-hand corner.

Other versions of Windows have slightly different paths to Data Sources (ODBC). There is also a helpful tutorial at <http://support.gfi.com/manuals/en/me56/tosetupnodbcdatasource.htm> for a good set of instructions.

You will need to set up the following data sources:

“students”, with a link to the path of the file “student.mdb”, for example, “c:\databases\student.mdb”

“games”, with a link to the path of the file “rad\_games.mdb”, for example, “c:\databases\rad\_games.mdb”.

# Appendix 3:

## Software for Chapters 9 and 10

---

In order to run and develop the exercises in Chapters 9 and 10 a VoiceXML browser is required. The IBM WebSphere VoiceServer SDK is recommended, but other platforms listed in Appendix 7 are also appropriate.

Software is also required to create a dynamic link to a database. Apache Tomcat 4.0 and Java(TM) 2 SDK, Version 1.3.1, were used to set up and develop the exercises in Chapters 9, 10 and 11. Other web servers and scripting languages may also be used.

### Downloading the IBM WebSphere VoiceServer SDK

The WebSphere VoiceServer SDK can be obtained from the following site: [http://www-3.ibm.com/software/pervasive/voice\\_toolkit/](http://www-3.ibm.com/software/pervasive/voice_toolkit/)

The IBM WebSphere Voice Server SDK includes a VoiceXML speech browser, speech recogniser, speech synthesiser, sample applications, and other tools for developing and testing VoiceXML applications. A comprehensive manual *VoiceXML Programmer's Guide* as well as the VoiceXML 1.0 Specification are provided as Adobe Acrobat files. The SDK runs in command mode and can be used independently of other software.

The following is an indication of the main requirements for the SDK:

#### Hardware

Intel® Pentium® 366MHz processor with minimum of 128MB RAM and 290 MB disk space.

A Microsoft® Windows® 2000 compatible 16-bit, full-duplex sound card (with a microphone input jack) with good recording quality.

A quality microphone.

#### Software

Microsoft Windows 2000 Professional or Server Service Pack 2. (*Note:* The SDK will work with Windows XP, although Windows XP is not officially supported.)

Sun Java Runtime Environment (Sun JRE) 1.3.1 (included in the package, but must be installed prior to the IBM WebSphere Voice Server SDK software).

Networking (e.g., an IP network) must be enabled.



Adobe Acrobat Reader, Version 5.0 (included in the package) or later.  
HTTP 1.1 Client, if desired.

Full installation instructions are provided with the download.

## ***Starting the VoiceXML Browser***

To start the VoiceXML browser, use the following command in the DOS command window:

```
%IBMVS%\bin\batchfile URL
```

IBMVS is the environment variable containing the location of the IBM WebSphere Voice Server SDK installation destination directory.

Batchfile is one of the following:

**vsaudio.** Generates spoken output and accepts spoken and simulated DTMF input (the language engine needs to be stipulated, e.g., vsaudio\_en\_US for the US language model).

**vstext.** Generates text output and accepts text or simulated DTMF input (the language engine needs to be stipulated, e.g., vstext\_en\_US for the US language model).

URL is the initial URL for your application.

Example:

```
"%IBMVS%\bin\vsaudio" myfile.vxml
```

In order to avoid lengthy path names for files on the computer's hard drive, it is best to change directory to the directory in which the files reside, e.g.,

```
cd c:\voicexml
```

## ***Voice Toolkit for WebSphere Studio***

The Voice Toolkit for WebSphere Studio works with the WebSphere Voice Server SDK for application development and testing, and with the IBM Reusable Dialog Components for adding VoiceXML components to voice applications. The Voice Toolkit is an integrated development environment for building VoiceXML applications and includes a VoiceXML editor and debugger, a grammar editor and test tool, pronunciation builder, built-in recorder, and VoiceXML reusable dialog components. The hardware and software requirements for the Voice Toolkit are listed at the web site. The Voice Toolkit is available as a free download. However, an important factor is that the Voice Toolkit is integrated into IBM's WebSphere Studio environment and requires either IBM WebSphere Studio Site Developer 5.0 or IBM WebSphere Studio Application Developer 5.0 to be installed. Site Developer and Application Developer are available with a free 60-day license, or professors can sign up for the IBM Scholars Program for WebSphere soft-

ware at the following site:

<http://www.ibm.com/software/info/university/products/websphere/>

## Downloading and Installing Apache Tomcat

Apache Tomcat can be obtained from the following site:

<http://jakarta.apache.org/tomcat/>

Apache Tomcat 4.0 is the version that was used for the files in Chapters 10 and 11.

## Java Server Pages (JSP)

The version of JSP that was used for the files in Chapters 10 and 11 was the Java(TM) 2 SDK, Standard Edition, Version 1.3.1, available at:

<http://java.sun.com/j2se/1.3/download.html>

Other versions of the SDK, such as 1.4, did not appear to work with Apache Tomcat 4.0.

# Appendix 4: Software for Chapter 11. IBM Multimodal Tools and Microsoft.NET Speech SDK

---

## IBM Multimodal Tools

The IBM Multimodal Tools are required to run the XHTML+Voice programs in Chapter 11. The Multimodal Tools can be obtained as a free download from: <http://www-3.ibm.com/software/pervasive/multimodal/>

The hardware and software requirements for the Multimodal Tools are listed at the web site. As with the Voice Toolkit (Appendix 3), the Multimodal Tools require that either IBM WebSphere Studio Site Developer 5.0 or IBM WebSphere Studio Application Developer 5.0 is installed. Site Developer and Application Developer are available with a free 60-day license, or professors can sign up for the IBM Scholars Program for WebSphere software at the following site: <http://www.ibm.com/software/info/university/products/websphere/>

## Microsoft .NET Speech SDK

The Microsoft .NET Speech SDK is required to run the SALT programs in Chapter 11. (Other environments for running SALT programs were described briefly in Chapter 11.) The SDK is available as a download from: <http://www.microsoft.com/speech/>

System requirements are listed at the site. The SDK is integrated into the Microsoft Visual Studio .NET 2003 environment which is required for the development and integration of SALT programs such as speech-enabled ASP.NET applications.

# Appendix 5: Spoken Dialogue Technology: Projects and Links

---

The following are links to a number of projects and sites involved in research and development in spoken dialogue technology.

**Table A5.1.** Links for spoken dialogue technology

---

## Companies

Apple Speech Recognition	<a href="http://www.apple.com/macosx/jaguar/speech.html">http://www.apple.com/macosx/jaguar/speech.html</a>
Dialogue Understanding Research Group (NTT, Japan)	<a href="http://www.brl.ntt.co.jp/cs/dug/index.html">http://www.brl.ntt.co.jp/cs/dug/index.html</a>
IBM Pervasive Computing	<a href="http://www-3.ibm.com/software/pervasive/multimodal/">http://www-3.ibm.com/software/pervasive/multimodal/</a>
Larson Technical Services – VoiceXML resources	<a href="http://www.larson-tech.com/bookres.htm">http://www.larson-tech.com/bookres.htm</a>
Microsoft Research	<a href="http://www.research.microsoft.com/research/projects/">http://www.research.microsoft.com/research/projects/</a>
Nuance	<a href="http://www.nuance.com">http://www.nuance.com</a>
Pipebeach	<a href="http://www.pipebeach.com/">http://www.pipebeach.com/</a>
Scansoft – leading supplier of imaging, speech and language solutions	<a href="http://www.scansoft.com">http://www.scansoft.com</a>
Telera (VoiceXML community)	<a href="http://www.telera.com/devxchange.html">http://www.telera.com/devxchange.html</a>
Vocalis – Voice- driven Business Solutions	<a href="http://www.vocalis.com/">http://www.vocalis.com/</a>
Voice Web Solutions – Development Tools for VoiceXML and SALT	<a href="http://www.voicewebsolutions.net">http://www.voicewebsolutions.net</a>

## Universities and Research Centres

Center for PersonKommunikation (CPK), Aalborg, Denmark	<a href="http://cpk.auc.dk/smc/index.html">http://cpk.auc.dk/smc/index.html</a>
---	---

**Table A5.1.** *Continued*

Centre for Communication Interface Research (CCIR), University of Edinburgh	<a href="http://www.ccir.ed.ac.uk/">http://www.ccir.ed.ac.uk/</a>
Centre for Language Technology (Macquarie University, Australia)	<a href="http://www.clt.mq.edu.au/Information/Overview.html">http://www.clt.mq.edu.au/Information/Overview.html</a>
Conversational Interaction and Spoken Dialogue Research Group, University of Rochester	<a href="http://www.cs.rochester.edu/research/cisd/">http://www.cs.rochester.edu/research/cisd/</a>
CSLR Home Page (Center for Spoken Language Research, University of Colorado)	<a href="http://cslr.colorado.edu">http://cslr.colorado.edu</a>
CSLU Home Page (Center for Spoken Language Understanding, Oregon)	<a href="http://cslu.cse.ogi.edu/">http://cslu.cse.ogi.edu/</a>
Department of Language and Speech (University of Nijmegen, Netherlands)	<a href="http://lands.let.kun.nl/">http://lands.let.kun.nl/</a>
GoDAG (University of Göteborg, Sweden)	<a href="http://www.ling.gu.se/grupper/GoDAG/">http://www.ling.gu.se/grupper/GoDAG/</a>
HCRC Dialogue Group (University of Edinburgh)	<a href="http://www.hcrc.ed.ac.uk/Site/DIALOGUE.html">http://www.hcrc.ed.ac.uk/Site/DIALOGUE.html</a>
KTH – Centre for Speech Technology, Stockholm, Sweden	<a href="http://www.speech.kth.se/ctt/">http://www.speech.kth.se/ctt/</a>
LIMSI: Projects on spoken language (France)	<a href="http://www.limsi.fr/Recherche/TLP/projects.html">http://www.limsi.fr/Recherche/TLP/projects.html</a>
Natural Dialogue Group (CSLU)	<a href="http://www.cslu.ogi.edu/ndg/">http://www.cslu.ogi.edu/ndg/</a>
Natural Interactive Systems Laboratory (NIS), Odense University, Denmark	<a href="http://www.nis.sdu.dk/">http://www.nis.sdu.dk/</a>
Natural Language Processing Group (University of Sheffield)	<a href="http://nlp.shef.ac.uk/">http://nlp.shef.ac.uk/</a>
NLPLAB (Linköping University, Sweden)	<a href="http://www.ida.liu.se/labs/nlplab/">http://www.ida.liu.se/labs/nlplab/</a>

**Table A5.1.** *Continued*

Speech Media Processing Group (University of Kyoto, Japan)	<a href="http://winnie.kuis.kyoto-u.ac.jp/home-e.html">http://winnie.kuis.kyoto-u.ac.jp/home-e.html</a>
Pattern Information Processing Group (Kyoto Institute of Technology, Japan)	<a href="http://www-vox.dj.kit.ac.jp/index-e.html">http://www-vox.dj.kit.ac.jp/index-e.html</a>
Speech Media Laboratory (University of Kyoto, Japan)	<a href="http://www.ar.media.kyoto-u.ac.jp/home-e.html">http://www.ar.media.kyoto-u.ac.jp/home-e.html</a>
Speech-based and Pervasive Interaction Group (Tampere, Finland)	<a href="http://www.cs.uta.fi/research/hci/spi/index.html">http://www.cs.uta.fi/research/hci/spi/index.html</a>
Spoken Language Systems Group (MIT)	<a href="http://www.sls.lcs.mit.edu/sls/sls-green-nospec.html">http://www.sls.lcs.mit.edu/sls/sls-green-nospec.html</a>
Tokyo Institute of Technology: Furui Laboratory	<a href="http://www.furui.cs.titech.ac.jp/english/index.html">http://www.furui.cs.titech.ac.jp/english/index.html</a>
Tutorial Dialogue Group (University of Edinburgh)	<a href="http://www.cogsci.ed.ac.uk/~jmoore/tutoring/links.html">http://www.cogsci.ed.ac.uk/~jmoore/tutoring/links.html</a>
University of Pittsburgh	<a href="http://www.cs.pitt.edu/~litman/">http://www.cs.pitt.edu/~litman/</a>
<b>Projects and Systems</b>	
Amities – Multilingual dialogue systems (University of Sheffield)	<a href="http://www.dcs.shef.ac.uk/nlp/amities/">http://www.dcs.shef.ac.uk/nlp/amities/</a>
August: A Swedish multimodal spoken dialogue system	<a href="http://www.speech.kth.se/august/">http://www.speech.kth.se/august/</a>
BusLine (CMU)	<a href="http://www-2.cs.cmu.edu/~aria/BusLine.html">http://www-2.cs.cmu.edu/~aria/BusLine.html</a>
CMU Communicator	<a href="http://www.speech.cs.cmu.edu/Communicator/">http://www.speech.cs.cmu.edu/Communicator/</a>
COMIC – Multimodal interaction (University of Edinburgh)	<a href="http://www.hcrc.ed.ac.uk/comic/">http://www.hcrc.ed.ac.uk/comic/</a>
Conversational Architectures Project	<a href="http://research.microsoft.com/adapt/conversation/">http://research.microsoft.com/adapt/conversation/</a>
CU Communicator	<a href="http://communicator.colorado.edu/">http://communicator.colorado.edu/</a>
DiaLeague – Forum for the evaluation of dialogue systems	<a href="http://dialeague.csl.sony.co.jp/">http://dialeague.csl.sony.co.jp/</a>
DINEX – Boston restaurant guide	<a href="http://www.sls.lcs.mit.edu/sls/applications/dinex.shtml">http://www.sls.lcs.mit.edu/sls/applications/dinex.shtml</a>
DIPPER	<a href="http://www.ltg.ed.ac.uk/dipper">http://www.ltg.ed.ac.uk/dipper</a>
DISC Best Practice Guide	<a href="http://www.disc2.dk/">http://www.disc2.dk/</a>

**Table A5.1.** *Continued*


---

DUG-1 spoken dialogue system (NTT, Japan)	<a href="http://www.brl.ntt.co.jp/cs/dug/dug1/">http://www.brl.ntt.co.jp/cs/dug/dug1/</a>
GALAXY (MIT)	<a href="http://www.sls.lcs.mit.edu/GALAXY.html">http://www.sls.lcs.mit.edu/GALAXY.html</a>
How May I Help You? (AT&T)	<a href="http://www.research.att.com/~algor/hmihy/">http://www.research.att.com/~algor/hmihy/</a>
Japanese dialogue system for an office robot	<a href="http://www-csli.stanford.edu/semlab/juno/">http://www-csli.stanford.edu/semlab/juno/</a>
Julietta Research Group (Seville, Spain)	<a href="http://fing.cica.es/">http://fing.cica.es/</a>
Jupiter (MIT)	<a href="http://www.sls.lcs.mit.edu/sls/whatwedo/applications/jupiter.html">http://www.sls.lcs.mit.edu/sls/whatwedo/applications/jupiter.html</a>
Language and Speech Technology programme (Netherlands)	<a href="http://odur.let.rug.nl:4321/">http://odur.let.rug.nl:4321/</a>
NASA spoken and multimodal dialogue systems	<a href="http://www.riacs.edu/research/detail/RIALIST_ver4/projects.htm">http://www.riacs.edu/research/detail/RIALIST_ver4/projects.htm</a>
Natural Language Research Group (Technical University of Catalonia, Spain)	<a href="http://www.lsi.upc.es/~nlp/">http://www.lsi.upc.es/~nlp/</a>
OLGA – Multimodal user interfaces and animated characters (Sweden)	<a href="http://www.nada.kth.se/~osu/olga/e_index.html">http://www.nada.kth.se/~osu/olga/e_index.html</a>
OVIS – Public transport system (Netherlands)	<a href="http://lands.let.kun.nl/TSPublic/strik/ovis.html">http://lands.let.kun.nl/TSPublic/strik/ovis.html</a>
SMARTKOM – Multimodal dialogue (Germany)	<a href="http://www.smartkom.org/start_en.html">http://www.smartkom.org/start_en.html</a>
Spoken Dialogue for Intelligent Tutoring Systems (University of Pittsburgh)	<a href="http://www2.cs.pitt.edu/~litman/why2-pubs.html">http://www2.cs.pitt.edu/~litman/why2-pubs.html</a>
Swedish Dialogue Systems	<a href="http://www.ida.liu.se/~nlplab/sds/">http://www.ida.liu.se/~nlplab/sds/</a>
TRAINS Project Home Page (University of Rochester)	<a href="http://www.cs.rochester.edu/research/trains/">http://www.cs.rochester.edu/research/trains/</a>
TRINDI	<a href="http://www.ling.gu.se/projekt/trindi/">http://www.ling.gu.se/projekt/trindi/</a>
TRIPS (University of Rochester)	<a href="http://www.cs.rochester.edu/research/cisd/projects/trips/">http://www.cs.rochester.edu/research/cisd/projects/trips/</a>
Universal Speech Interface (CMU)	<a href="http://www-2.cs.cmu.edu/~usi/">http://www-2.cs.cmu.edu/~usi/</a>
Verbmobil project (Germany)	<a href="http://verbmobil.dfki.de/overview-us.html">http://verbmobil.dfki.de/overview-us.html</a>

**Table A5.1.** *Continued*


---

Voyager – Tourist and travel information (MIT)	<a href="http://www.sls.lcs.mit.edu/sls/applications/voyager.shtml">http://www.sls.lcs.mit.edu/sls/applications/voyager.shtml</a>
Waxholm dialog project (Sweden)	<a href="http://www.speech.kth.se/waxholm/waxholm2.html">http://www.speech.kth.se/waxholm/waxholm2.html</a>
WITAS – Multimodal conversational interfaces	<a href="http://www-csli.stanford.edu/semlab/witas/">http://www-csli.stanford.edu/semlab/witas/</a>
<b>Resources</b>	
Dialogue Diversity Corpus	<a href="http://www-rcf.usc.edu/~billmann/diversity">http://www-rcf.usc.edu/~billmann/diversity</a>
Dialogue systems links (Dan Bohus)	<a href="http://www-2.cs.cmu.edu/~dbohus/SDS/">http://www-2.cs.cmu.edu/~dbohus/SDS/</a>
Dialogue systems links (Emiel Kraemer)	<a href="http://fdlwww.kub.nl/~kraemer/usimodule-C4-2001.htm">http://fdlwww.kub.nl/~kraemer/usimodule-C4-2001.htm</a>
Dialogue systems links (Gabriel Skantze)	<a href="http://www.speech.kth.se/~gabriel/speech.html">http://www.speech.kth.se/~gabriel/speech.html</a>
Dialogue systems links (Jens-Uwe Möller)	<a href="http://nats-www.informatik.uni-hamburg.de/~jum/research/dialog/sys.html">http://nats-www.informatik.uni-hamburg.de/~jum/research/dialog/sys.html</a>
Dialogue systems links (Steffan Larsson)	<a href="http://www.ling.gu.se/~sl/dialogue_links.html">http://www.ling.gu.se/~sl/dialogue_links.html</a>
List of spoken dialogue systems currently in operation	<a href="http://www.disc2.dk/tools/opSLDSs.html">http://www.disc2.dk/tools/opSLDSs.html</a>
SALT Forum	<a href="http://www.saltforum.org/">http://www.saltforum.org/</a>
SIGDIAL – Special interest group of ACL for dialogue and discourse	<a href="http://www.sigdial.org/">http://www.sigdial.org/</a>
Spoken Dialogue systems links	<a href="http://liceu.uab.es/~joaquim/speech_technology/tecnol_parla/dialogue/refs_dialeq.html">http://liceu.uab.es/~joaquim/speech_technology/tecnol_parla/dialogue/refs_dialeq.html</a>
VoiceXML Forum	<a href="http://www.voicexml.org">http://www.voicexml.org</a>
W3C “Voice Browser” activity	<a href="http://www.w3.org/Voice/">http://www.w3.org/Voice/</a>

---



# Appendix 6:

## Spoken Dialogue Technology: Toolkits and Resources

---

### Toolkits and Resources for Spoken Dialogue Technology

**Table A6.1.** Spoken dialogue technology: Toolkits and resources

---

CODIAL – Tutorial and tool in support of cooperative dialogue design for spoken language dialogue systems	<a href="http://www.disc2.dk/tools/codial/">http://www.disc2.dk/tools/codial/</a>
CSLU toolkit	<a href="http://cslu.cse.ogi.edu/toolkit">http://cslu.cse.ogi.edu/toolkit</a>
CU Communicator	<a href="http://communicator.colorado.edu/">http://communicator.colorado.edu/</a>
DRI Shared tools and resources	<a href="http://www.georgetown.edu/faculty/luperfos/Discourse-Treebank/tools-and-resources.html">http://www.georgetown.edu/faculty/luperfos/Discourse-Treebank/tools-and-resources.html</a>
IBM WebSphere Multimodal Tools	<a href="http://www-3.ibm.com/software/pervasive/multimodal/">http://www-3.ibm.com/software/pervasive/multimodal/</a>
IBM WebSphere VoiceServer SDK	<a href="http://www-3.ibm.com/software/pervasive/products/voice/voice_server_sdk.shtml">http://www-3.ibm.com/software/pervasive/products/voice/voice_server_sdk.shtml</a>
MATE workbench – A software tool set for multilevel and cross-level annotation of corpora and extraction of information about annotated corpora	<a href="http://mate.nis.sdu.dk/">http://mate.nis.sdu.dk/</a>
Microsoft Speech Technologies	<a href="http://www.microsoft.com/speech/">http://www.microsoft.com/speech/</a>
Mobile Conversay Software Development Kit (SDK)	<a href="http://www.conversa.com">http://www.conversa.com</a>
Natural Language Speech Assistant (NLSA) (Unisys Corporation)	<a href="http://www.unisys.com/">http://www.unisys.com/</a>
NUANCE Developers Toolkit	<a href="http://www.nuance.com">http://www.nuance.com</a>
SpeechWorks OpenVXI 2.0.1	<a href="http://www.speech.cs.cmu.edu/openvxi">http://www.speech.cs.cmu.edu/openvxi</a>
SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces	<a href="http://guir.berkeley.edu/projects/suede/">http://guir.berkeley.edu/projects/suede/</a>
Telera AppBuilder (VoiceXML tool)	<a href="http://www.telera.com/appbuilder.html">http://www.telera.com/appbuilder.html</a>
TRINDIKIT	<a href="http://www.ling.gu.se/projekt/trindi//trindikit/">http://www.ling.gu.se/projekt/trindi//trindikit/</a>

---

---

## Web Sites for Hosted VoiceXML Applications

**Table A6.2.** Web sites for hosted VoiceXML Applications

---

BeVocal Cafe	<a href="http://cafe.bevocal.com/">http://cafe.bevocal.com/</a>
HeyAnita	<a href="http://freespeech.heyanita.com/">http://freespeech.heyanita.com/</a>
Tellime Studio	<a href="http://studio.tellme.com/">http://studio.tellme.com/</a>
VoiceGenie	<a href="http://developer.voicegenie.com/">http://developer.voicegenie.com/</a>
Voxeo	<a href="http://community.voxeo.com/">http://community.voxeo.com/</a>
Voxpilot	<a href="http://www.voxpilot.com/">http://www.voxpilot.com/</a>

---

# References

---

- Aberdeen J, Bayer S, Caskey S, Damianos L, Goldschen A, Hirschman L, Loehr D, Trappe H (1999). DARPA Communicator Program tackles conversational interface challenges. *The Edge* (Intelligent Human-Computer Interface Issue) 3:4. MITRE Publications. Available at:  
[http://www.mitre.org/news/the\\_edge/december\\_99/third.html](http://www.mitre.org/news/the_edge/december_99/third.html)
- Alexandersson J, Buschbeck-Wolf B, Fujinami T, Kipp T, Koch S, Maier E, Reithinger N, Schmitz B, Siegel M (1998). *Dialogue Acts in Verbmobil – 2*, 2nd ed. Verbmobil-Report Nr. 226. DFKI GmbH, Saarbrücken.
- Allen JF (1983). Recognising intentions from natural language utterances. In: Brady M, Berwick RC (eds.), *Computational Models of Discourse*. MIT Press, Cambridge, MA.
- Allen JF (1995). *Natural Language Processing*, 2nd ed. Benjamin Cummings, Redwood City, CA.
- Allen JF, Core M (1997). Draft of DAMSL: Dialog act markup in several layers. Available at:  
<http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/RevisedManual.html>
- Allen JF, Perrault CR (1980). Analysing intentions in utterances. *Artificial Intelligence* 15:143–178.
- Allen JF, Ferguson G, Miller BW, Ringger EK, Zollo TS (2000). Dialogue systems: From theory to practice in TRAINS-96. In: Dale R, Moisl H, Somers H (eds.), *Handbook of Natural Language Processing*. Marcel Dekker, New York, pp. 347–376.
- Allen JF, Byron DK, Dzikovska M, Ferguson G, Galescu L, Stent A (2001a). Towards conversational human-computer interaction. *AI Magazine* 22(4):27–38.
- Allen JF, Ferguson G, Stent A (2001b). An architecture for more realistic conversational systems. *Proceedings of Intelligent User Interfaces (IUI-01)*, Santa Fe, NM.
- Allwood J (1976). *Linguistic Communication as Action and Cooperation*. Gothenburg Monographs in Linguistics, Vol. 2. Department of Linguistics, University of Göteborg.

- Andersson EA, Breitenbach S, Burd T, Chidambaram N, Houle P, Newsome D, Tang X, Zhu X (2001). *Early Adopter VoiceXML*. Wrox Press, Birmingham, UK.
- André E (2000). The generation of multimedia presentations. In: Dale R, Moisl H, Somers H (eds.), *Handbook of Natural Language Processing*. Marcel Dekker, New York, pp. 305–327.
- André E, Rist T (1993). The design of illustrated documents as a planning task. In: Maybury M (ed.), *Intelligent Multimedia Interfaces*. AAAI Press, Menlo Park, CA.
- André E, Rist T (1996). Coping with temporal constraints in multimedia presentation planning. *Proceedings of the 13th National Conference of the American Association on Artificial Intelligence (AAAI'96)*, Portland, OR, pp. 142–147.
- André E, Müller J, Rist T (1996). The PPP Persona: A multi-purpose animated presentation agent. In: Catarci T, Costabile MF, Levialdi S, Santucci G (eds.), *Advanced Visual Interfaces*. ACM Press, New York, pp. 245–247.
- Androutopoulos I, Ritchie G (2000). Database interfaces. In: Dale R, Moisl H, Somers H (eds.), *Handbook of Natural Language Processing*. Marcel Dekker, New York.
- Androutopoulos I, Aretoulaki M (2003). Natural language interaction. In: Mitkov R (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford, UK, pp. 629–649.
- Aretoulaki M, Ludwig B (1999). Automaton descriptions and theorem-proving: A marriage made in heaven? *Proceedings of the IJCAI'99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Stockholm, Sweden.
- Argyle M (1975). *Bodily Communication*. Methuen, London.
- ARPA (1994). *Proceedings of the Speech and Natural Language Workshop*. Morgan Kaufmann, San Mateo, CA.
- Atkinson JM, Drew P (1979). *Order in Court: The Organization of Verbal Interaction in Judicial Settings*. Macmillan, London.
- Aust H, Oerder M, Seide F, Steinbiss V (1995). The Philips automatic train timetable information system. *Speech Communication* 17:249–262.
- Austin JL (1962). *How to Do Things with Words*. Oxford University Press, Oxford, UK.
- Balentine B, Morgan DP, Meisel WS (2001). *How to Build a Speech Recognition Application: A Style Guide for Telephony Dialogues*, 2nd ed. Enterprise Integration Group, San Ramon, CA.
- Bales RF (1950). *Interaction Process Analysis*. Addison-Wesley, Cambridge, MA.
- Batacharia B, Levy D, Catizone R, Krotov A, Wilks Y (1997). CONVERSE: A conversational companion. *Proceedings of the 1st International Workshop on Human-Computer Conversation*, Bellagio, Italy. (Reprinted in: Wilks Y (1999) (ed.), *Machine Conversations*. Kluwer Academic, Dordrecht.)
- Benoit C, Martin JC, Pelachaud C, Schomaker L, Suhm B (2000). Audio-visual and multimodal speech-based systems. In: Gibbon D, Mertins I, Moore R (eds.), *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. Kluwer Academic, Dordrecht, pp. 102–203.

- Bergsten H (2002). *JavaServer Pages*, 2nd ed. O'Reilly, Sebastopol, CA.
- Bernsen NO (2003). On-line user modelling in a mobile spoken dialogue system. *Proceedings of the 8th International Conference on Speech Communication and Technology (Eurospeech2003)*, Geneva, September, pp. 737–740.
- Bernsen NO, Dybkjær L (1997). The DISC project. *ELRA Newsletter* 2(2):6–8. Available at: <http://www.disc2.dk/publications/other/INFO-1-Elra.html>
- Bernsen NO, Dybkjær H, Dybkjær L (1998). *Designing Interactive Speech Systems: From First Ideas to User Testing*. Springer-Verlag, New York.
- Bernsen NO, Dybkjær L (2002). A multimodal virtual co-driver's problem with the driver. *Proceedings of IDS02*, Kloster Irsee, Germany, June.
- Bilmes J (1988). The concept of preference in conversation analysis. *Language in Society* 17(2):161–182.
- Bobrow DG (1968). Natural language input for a computer problem-solving system. In: Minsky M (ed.), *Semantic Information Processing*. MIT Press, Cambridge, MA.
- Bobrow DG, Kaplan RM, Kay M, Norman DA, Thompson H, Winograd T (1977). GUS: A frame-driven dialogue system. *Artificial Intelligence* 8:155–173.
- Boros M, Eckert W, Gallwitz F, Görz G, Hanrieder G, Niemann H (1996). Towards understanding spontaneous speech: Word accuracy vs. concept accuracy. *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP96)*, Philadelphia, PA, October, pp. 1005–1008.
- Bos J, Klein E, Lemon O, Oka T (2003). DIPPER: Description and formalisation of an information-state update dialogue system architecture. *Proceedings of the 4th SIGDial Workshop on Dialogue and Discourse*, Sapporo, Japan, 5–6 July.
- Bouwman A, Sturm J, Boves L (1999). Incorporating confidence measures in the Dutch train timetable information system developed in the Arise project. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, Phoenix, AZ, pp. 493–496.
- Brennan SE, Hulstén E (1995). Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems* 8:143–151.
- Brown G, Yule G (1983). *Discourse Analysis*. Cambridge University Press, Cambridge, UK.
- Bunt HC (1979). Conversational principles in question–answer dialogues. In: Krallmann D (ed.), *Zur Theorie der Frage*. Narr Verlag, Essen, pp. 119–141.
- Bunt HC, Black B (2000a) (eds.). *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*. John Benjamins, Amsterdam.
- Bunt HC, Black B (2000b). The ABC of computational pragmatics. In: Bunt HC, Black B (eds.), *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*. John Benjamins, Amsterdam, pp. 1–46.
- Burnett DC (2000) (ed.). *SpeechObjects Specification V1.0. W3C Note* November. Available at: <http://www.w3.org/TR/speechobjects>
- Button G, Coulter J, Lee JRE, Sharrock W (1995). *Computers, Minds and Conduct*. Polity Press, Cambridge, UK.
- Cahn JE, Brennan SE (1999). A psychological model of grounding and repair in dialog. *Proceedings of the AAAI Fall Symposium on Psychological Models of*

- Communication in Collaborative Systems*. North Falmouth, MA. AAAI Press, Menlo Park, CA, pp. 25–33.
- Carberry MS (1986). The use of inferred knowledge in understanding pragmatically ill-formed queries. In: Reilly R (ed.), *Communication Failure in Dialogue and Discourse*. North-Holland, Amsterdam.
- Carbonell J (1981). POLITICS. In: Schank RC, Riesbeck CK (eds.), *Inside Computer Understanding*. Lawrence Erlbaum, Hillsdale, NJ.
- Carletta J, Isard A, Isard S, Kowtko J, Doherty-Sneddon G, Anderson A (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics* 23:13–31.
- Carlson R, Granström B (1997). Speech synthesis. In: Hardcastle WJ, Laver J (eds.), *The Handbook of Phonetic Science*. Blackwell, Oxford, UK, pp. 768–788.
- Cassell J, Sullivan J, Prevost S, Churchill E (2000) (eds.). *Embodied Conversational Agents*. MIT Press, Cambridge, MA.
- Chin D (1989). KNOPE: Modeling what the user knows in UC. In: Kobsa A, Wahlster W (eds.), *User Models in Dialog Systems*. Springer-Verlag, London, pp. 74–107.
- Clark HH (1996). *Using Language*. Cambridge University Press, Cambridge, UK.
- Clark HH (2003). When to start speaking, when to stop, and how. Invited talk: *Proceedings of the ISCA Workshop on Error Handling in Spoken Dialogue Systems*, pp. 1–4.
- Clark HH, Schaefer EF (1989). Contributing to discourse. *Cognitive Science* 13:259–294.
- Cohen PR (1994). Models of dialogue, cognitive processing for vision and voice. In: Nagao M (ed.), *Proceedings of the Fourth NEC Research Symposium*. SIAM, Philadelphia, PA.
- Cohen PR, Johnston M, McGee D, Oviatt SL, Pittman J, Smith I, Chen L, Chow J (1997). Quickset: Multimodal interaction for distributed applications. *Proceedings of the Fifth ACM International Multimedia Conference*. ACM Press, New York, pp. 31–40.
- Cohen W (1996). Learning trees and rules with set-valued features. *Fourteenth Conference of the American Association of Artificial Intelligence*.
- Colby KM (1975). *Artificial Paranoia – A Computer Simulation of Paranoid Processes*. Pergamon Press, New York.
- Cole RA, Mariani J, Uskoreit H, Zaenen A, Zue V (1997) (eds.). *Survey of the State of the Art in Human Language Technology*. Cambridge University Press, Cambridge, UK. Available at: <http://cslu.cse.ogi.edu/HLTsurvey>
- Cole R, Massaro DW, de Villiers J, Rundle B, Shobaki K, Wouters J, Cohen M, Beskow J, Stone P, Connors P, Tarachow A, Solcher D (1999). New tools for interactive speech and language training: Using animated conversational agents in the classrooms of profoundly deaf children. *Proceedings of the ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education*, London, April.
- Connors P, Davis A, Fortier G, Gilley K, Rundle B, Soland C, Tarachow A (1999). Participatory design: Classroom applications and experiences. *Proceedings of the International Conference of Phonetic Sciences*, San Francisco, CA.

- Cooper R, Larsson S, Matheson C, Poesio M, Traum DR (1999). Coding Instructional Dialogue for Information States Trindi Project Deliverable D1.1. Available at: <http://www.ling.gu.se/projekt/trindi/publications.html>
- Cullingford R (1981). SAM. In: Schank RC, Riesbeck CK (eds.), *Inside Computer Understanding*. Lawrence Erlbaum, Hillsdale, NJ.
- Dahlbäck N, Jönsson A (1998). A coding manual for the Linköping dialogue model. Available at: <http://www.cs.umd.edu/users/traum/DSD/arne2.ps>
- DAMSL Dialog Annotation (2001). Available at: <http://www.cs.rochester.edu/research/cisd/resources/damsl/>
- Danieli M, Gerbino E (1995). Metrics for evaluating dialogue strategies in a spoken language system. *Working Notes of the AAAI Spring Symposium on Empirical Methods on Discourse Interpretation and Generation*. AAAI, Stanford, CA, pp. 34–39.
- DARPA (1992). *Proceedings of the Speech and Natural Language Workshop*. Morgan Kaufmann, San Mateo, CA.
- den Os E, Boves L, Lamel L, Baggia P (1999). Overview of the ARISE project. *Proceedings of the 6th International Conference on Speech Communication and Technology (Eurospeech99)*, Budapest, Hungary, September, pp. 1527–1530.
- Eckert W, Nöth E, Niemann H, Schukat-Talamazzan EG (1995). Real users behave weird: Experiences made collecting large human-machine dialog corpora. *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, Vigso, Denmark, pp. 193–196.
- Edgington M, Lowry A, Jackson P, Breen A, Minnis S (1996a). Overview of current Text-to-Speech techniques: Part I – Text and linguistic analysis. *BT Technical Journal* 14:68–83.
- Edgington M, Lowry A, Jackson P, Breen A, Minnis S (1996b). Overview of current Text-to-Speech techniques: Part II – Prosody and speech generation. *BT Technical Journal* 14:84–99.
- Feiner SK, McKeown KR (1991). Automating the generation of co-ordinated multimedia explanations. *IEEE Computer* 24(10):33–41.
- Fillmore CH (1968). The case for case. In: Bach E, Harms RT (eds.), *Universals in Linguistic Theory*. Holt, Rinehart and Winston, New York, pp. 1–88.
- Fraser N, Gilbert GN (1991). Simulating speech systems. *Computer Speech and Language* 5:81–99.
- Gardner-Bonneau D (1999) (ed.). *Human Factors and Voice Interactive Systems*. Kluwer Academic, Boston, MA.
- Giachin E, McGlashan S (1997). Spoken language dialogue systems. In: Young S, Bloothoof G (eds.), *Corpus-based Methods in Language and Speech Processing*. Kluwer Academic, Dordrecht, pp. 69–117.
- Gibbon D, Moore R, Winski R (1997) (eds.). *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, New York.
- Glass JR (1999). Challenges for spoken dialogue systems. *Proceedings of the IEEE ASRU Workshop*, Keystone, CO, December.
- Glass JR, Polifroni J, Seneff S, Zue V (2000). Data collection and performance evaluation of spoken dialogue systems: The MIT experience. *Proceedings of*

- the 6th International Conference on Spoken Language Processing (ICSLP2000)*, Beijing, China.
- Glass JR, Weinstein E (2001). SPEECHBUILDER: Facilitating spoken dialogue system development. *Proceedings of the 7th European Conference on Speech Communication and Technology*, Aalborg, Denmark, September.
- Goddeau D, Meng H, Polifroni J, Seneff S, Busayapongchai S (1996). A form-based dialogue manager for spoken language applications. *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP'96)*, Pittsburgh, PA, pp. 701–704.
- Graesser AC, VanLehn K, Rose C, Jordan P, Harter D (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine* 22:39–51.
- Green BF, Wolf AW, Chomsky C, Laughery KR (1963). BASEBALL: An automatic question–answerer. In: Feigenbaum EA, Feldman J (eds.), *Computers and Thought*. McGraw-Hill, New York.
- Grice HP (1975). Logic and conversation. In: Cole P, Morgan JL (eds.), *Syntax and Semantics*, Vol. 3: *Speech Acts*. Academic Press, New York.
- Grosz BJ (1978). Discourse. In: Walker D (ed.), *Understanding Spoken Language*. Elsevier North-Holland, New York.
- Grosz BJ, Sidner CL (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3):175–204.
- Hauptmann AG, Rudnicky AI (1988). Talking to computers: An empirical investigation. *International Journal of Man–Machine Studies* 28(6):583–604.
- Heeman PA, Allen JF (1994). Detecting and correcting speech repairs. *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94)*, Las Cruces, June, pp. 295–302.
- Heisterkamp P (2001). Linguatronic – Product level speech system for Mercedes-Benz cars. *Proceedings of the HLT 2001*. Morgan Kaufmann, San Francisco, CA.
- Heisterkamp P, McGlashan S (1996). Units of dialogue management: An example. *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP'96)*, Philadelphia, PA, pp. 200–203.
- Hirschberg J (2002). Communication and prosody: Functional aspects of prosody. *Speech Communication* 36:31–43.
- Hirschberg J, Litman DJ, Swerts M (1999). Prosodic cues to recognition errors. *Proceedings of the Automatic Speech Recognition and Understanding Workshop (ASRU99)*, Keystone, CO, December.
- Hirschman L (1994). The roles of language processing in a spoken language interface. In: Roe DG, Wilpon JG (eds.), *Voice Communication Between Humans and Machines*. National Academy Press, Washington, DC.
- Hosom JP, Cole RA, Fandy M, Colton D (1997). Speech recognition using neural networks at the Center for Spoken Language Understanding. Available at: <http://www.cslu.ogi.edu/toolkit/old/old/documentation/recog/recog.html>
- Huang X, Acero A, Hon HW (2001). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Prentice Hall, Upper Saddle River, NJ.



- Jefferson G (1989). Preliminary notes on a possible metric which provides for a “standard maximum” silence of approximately one second in conversation. In: Roger D, Bull P (eds.), *Conversation: An Interdisciplinary Perspective. Multilingual Matters*, Clevedon, UK.
- Johnston M, Cohen PR, McGee D, Oviatt SL, Pittman JA, Smith I (1997). Unification-based multimodal integration. *Proceedings of the International Conference on Computational Linguistics and the 35th Meeting of the Association for Computational Linguistics (COLING-ACL'97)*, Madrid, pp. 281–288.
- Johnston M, Bangalole S, Vasireddy G, Stent A, Eblen P, Walker MA, Whittaker S, Maloor P (2002). MATCH: An architecture for multimodal dialogue systems. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, PA, July, pp. 376–383.
- Jokinen K, Wilcock G (to appear). Confidence-based adaptivity and response generation in a spoken dialogue system. In: van Kuppevelt J, Smith R (eds.), *Current and New Directions in Discourse and Dialogue*. Kluwer Academic, Dordrecht.
- Jokinen K, Kerminen A, Kaipainen M, Jauhiainen T, Wilcock G, Turunen M, Hakulinen J, Kuusisto J, Lagus K (2002). Adaptive dialogue systems – Interaction with interact. *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue*, Philadelphia, PA, July 11–12.
- Jurafsky D (2004). Pragmatics and computational linguistics. In: Horn L, Ward G (eds.), *Handbook of Pragmatics*. Blackwell, Oxford, UK.
- Jurafsky D, Shriberg E, Biasca D (1997). Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Draft 13. University of Colorado, Boulder, CO. Institute of Cognitive Science Technical Report 97-02. Available at:  
<http://www.colorado.edu/ling/jurafsky/ws97/manual.august1.html>
- Jurafsky D, Martin JH (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, Upper Saddle River, NJ.
- Kamm C, Walker MA, Litman DJ (1999). Evaluating spoken language systems. *Proceedings of the American Voice Input/Output Society (AVIOS)*.
- Kamp H, Reyle U (1993). *From Discourse to Logic: Introduction to Model Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic, Dordrecht.
- Kaplan J (1983). Co-operative responses from a portable natural language database query system. In: Brady M, Berwick RC (eds.), *Computational Models of Discourse*. MIT Press, Cambridge, MA.
- Karsenty L (2002). Shifting the design philosophy of spoken dialogue systems: From invisible to transparent systems. *International Journal of Speech Technology* 5(2):147–157.
- Kelsey Group (2002). Speech technology revenues to reach \$2 billion by 2006 according to the Kelsey Group. Available at:  
<http://www.kelseygroup.com/press/pr020430.htm>
- Kerminen A, Jokinen K (2003). Distributed dialogue management in a blackboard architecture. *Proceedings of the EACL Workshop Dialogue Systems:*

- Interaction, Adaptation and Styles of Management*, Budapest, Hungary, pp. 55–66.
- Klabbers E, van Santen JPH (2003). Control and prediction of the impact of pitch modification on synthetic speech quality. *Proceedings of the 8th International Conference on Speech Communication and Technology (Eurospeech2003)*, Geneva, September, pp. 317–320.
- Komatani K, Adachi F, Ueno S, Kawahara T, Okuno HG (2003). Flexible spoken dialogue system based on user models and dynamic generation of VoiceXML scripts. *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, Sapporo, Japan, July 5–6, pp. 87–96.
- Kowtko J, Isard S, Doherty GM (1993). Conversational Games Within Dialogue. Research Paper HCRC/RP-31. Human Communication Research Centre, University of Edinburgh, Scotland, UK.
- Krahmer E, Swerts M, Theune T, Weegels ME (2001). Error detection in spoken human-machine interaction. *International Journal of Speech Technology* 4(1):19–30.
- Krahmer E, Swerts M, Theune T, Weegels ME (2002). The dual of denial: Two uses of disconfirmations in dialogue and their prosodic correlates. *Speech Communication*. 36(1):133–145.
- Kurzweil R (2001). *The Age of Spiritual Machines: How We Will Live, Work and Think in the New Age of Intelligent Machines*. Texere Publishing, NY.
- Lai J (2000) (ed.). Conversational interfaces. *Communications of the ACM* 43:9.
- Lai J, Yankelovich N (2002). Conversational speech interfaces. In: Sears A, Jacko J (eds.), *The Handbook of Human-Computer Interaction*. Lawrence Erlbaum, Hillsdale, NJ.
- Lambert T, Breen AP, Eggleton B, Cox SJ, Milner BP (2003). Unit selection in concatenative TTS synthesis systems based on Mel filter bank amplitudes and phonetic context. *Proceedings of the 8th International Conference on Speech Communication and Technology (Eurospeech2003)*, Geneva, September, pp. 273–276.
- Lamel L, Bennacef S, Bonneau-Maynard H, Rosset S, Gauvain JL (1995). Recent developments in spoken language systems for information retrieval. *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, Vigsø, Denmark, pp. 17–20.
- Larson JA (2000). Speech-enabled appliances. *Speech Technology Magazine*, November/December 2000. Available at: [http://www.speechtechmag.com/issues/5\\_6/cover/231-1.html](http://www.speechtechmag.com/issues/5_6/cover/231-1.html)
- Larson JA (2002). *VoiceXML: Introduction to Developing Speech Applications*. Prentice Hall, Upper Saddle River, NJ.
- Larsson S, Traum DR (2000). Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. *Natural Language Engineering Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering*, pp. 323–340.
- Larsson S (2002). Issue-based Dialogue Management. PhD thesis, Goteborg University.

- Leech G, Weisser M (2003). Pragmatics and dialogue. In: Mitkov R (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford, UK.
- Lehnert WG (1980). Question answering in natural language processing. In: Bolc L (ed.), *Natural Language Question Answering*. Hanser, Munich.
- Levin E, Pieraccini R, Eckert W (2000). A stochastic model of human-machine interaction for learning dialog strategies. *IEEE Transactions on Speech and Audio Processing* 8(1):11–22.
- Levinson SC (1983). *Pragmatics*. Cambridge University Press, Cambridge, UK.
- Litman DJ (2002). Adding spoken dialogue to a text-based tutorial dialogue system. *Proceedings of the ITS Workshop on Empirical Methods for Tutorial Dialogue Systems*, Biarritz, France, June.
- Litman DJ, Pan S (1999). Empirically evaluating an adaptable spoken dialogue system. *Proceedings of the 7th International Conference on User Modeling (UM'99)*, pp. 55–64.
- Litman DJ, Pan S (2002). Designing and evaluating an adaptive spoken dialogue system. *User Modeling and User-Adapted Interaction* 12(2/3):111–137.
- Litman DJ, Hirschberg J, Swerts M (2000). Predicting automatic speech recognition performance using prosodic cues. *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 218–225.
- Litman DJ, Walker MA, Kearns MS (1999). Automatic detection of poor speech recognition at the dialogue level. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL'99)*, College Park, MD, June, pp. 309–316.
- Mann W, Thompson S (1988). Rhetorical structure theory: Toward a functional theory of text organisation. *Text* 3:243–281.
- Markowitz JA (1996). *Using Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ.
- Matheson C, Poesio M, Traum DR (2001). PTT and EDIS. In: TRINDI Consortium, 2001. The TRINDI book (Draft). Available at: <http://www.ling.gu.se/projekt/trindi/publications.html>
- Maudlin M (1994). Chatterbots, tinymuds and the Turing test: Entering the Loebner prize competition. *Proceedings of AAAI-94*, August, Seattle.
- McKeown K (1985). *Text Generation*. Cambridge University Press, Cambridge, UK.
- McKevitt P (1994–1996) (ed.). The integration of natural language and vision processing. Special issues of *Artificial Intelligence Review* 8(2–3, 4–5), 9(2–3, 5–6), 10(1–2, 3–4). Also available as McKevitt P (1995/1996). *The Integration of Natural Language and Vision Processing*, Vols. 1–4. Kluwer Academic, Dordrecht.
- McRoy SW (1998). Preface – Detecting, repairing and preventing human-machine miscommunication. *International Journal of Human-Computer Studies*, 48:547–552.
- McTear MF (1987). *The Articulate Computer*. Blackwell, Oxford, UK.

- McTear MF (2002). Spoken dialogue technology: Enabling the conversational interface. *ACM Computing Surveys* 34(1):90–169.
- Milward D, Thomas J, Amores G, Bos J, Boye J, Cooper R, Ericsson S, Gorrell G, Hjelm D, Holt A, Knight S, Larsson S, Lewin I, Quesada JF (2001). A D’Homme Demonstrator in English, Swedish and Spanish. Dialogues in the Home Machine Environment: IST-2000-26280 Deliverable D2.2, October, 2001. Available at: <http://www.ling.gu.se/projekt/dhomme>
- Minker W, Haiber U, Heisterkamp P (2002). Intelligent dialogue strategy for accessing infotainment applications in mobile environments. *Proceedings of IDS02*, Kloster Irsee, Germany, June.
- Mitkov R (2003). (ed.). *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford, UK.
- Moore RC (1994). Integration of speech with natural language understanding. In: Roe DG, Wilpon JG (eds.), *Voice Communication Between Humans and Machines*. National Academy Press, Washington, DC, pp. 254–271.
- Murphy A (2001). Interactive Learning Using Speech: The use of speech as a tool for primary school children. MSc Dissertation Computing and Information Systems, University of Ulster.
- Nakatani C, Traum DR (1999). Coding Discourse Structure in Dialogue (Version 1.0). Technical Report UMIACS-TR-99-03, University of Maryland. Available at: <http://www.cs.umd.edu/users/traum/Papers/papers.html>
- Neal JG, Shapiro SC (1991). Intelligent multimedia interface technology. In: Sullivan J, Tyler S (eds.), *Intelligent User Interfaces*. ACM Press, New York, pp. 11–43.
- Nickell J (2002). Are you being served? *Technology Review*, March 15. Available at: [http://www.technologyreview.com/articles/print\\_version/nickell031502.asp](http://www.technologyreview.com/articles/print_version/nickell031502.asp)
- Nijholt A, Heylen D (2002). Multimodal communication in inhabited virtual environments. *International Journal of Speech Technology* 5:343–354.
- Nöth E, Batliner A, Warnke V, Haas J, Boros M, Buckow J, Huber R, Gallwitz F, Nutt M, Niemann H (2002). On the use of prosody in automatic dialogue understanding. *Speech Communication* 36:45–62.
- Novick DG (2002). (ed.). Special section on natural language interfaces. *International Journal of Speech Technology* 5(2).
- O’Neill IM, McTear MF (2002). A pragmatic confirmation mechanism for an object-based spoken dialogue manager. *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP’2002)*, Vol. 3, Denver, CO, September, pp. 2045–2048.
- O’Neill IM, Hanna P, Liu X, McTear MF (2003). The Queen’s Communicator: An object-oriented dialogue manager. *Proceedings of the 8th International Conference on Speech Communication and Technology (Eurospeech2003)*, Geneva, September, pp. 593–596.
- Ousterhout JK (1994). *TCL and the TK Toolkit*. Addison-Wesley, Boston, MA.
- Oviatt SL (1997). Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction* (Special Issue on Multimodal Interfaces) 12:93–129.

- Oviatt SL (1999). Mutual disambiguation of recognition errors in a multimodal architecture. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'99)*. ACM Press, New York, pp. 576–583.
- Oviatt SL (2000). Taming recognition errors with a multimodal interface. *Communications of the ACM* 43(9):45–51.
- Oviatt SL (2002). Multimodal interfaces. In: Jacko J, Sears A (eds.), *Handbook of Human-Computer Interaction*. Lawrence Erlbaum, Hillsdale, NJ.
- Paek T, Horvitz E (1999). Uncertainty, utility, and misunderstanding. *AAAI Fall Symposium on Psychological Models of Communication*, North Falmouth, MA, November 5–7, pp. 85–92.
- Paek T, Horvitz E (2000). Conversation as action under uncertainty. *Sixteenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, San Francisco, CA, pp. 455–464.
- Page JH, Breen AP (1996). The Laureate text-to-speech system – Architecture and applications. *BT Technology Journal* 14(1):57–67.
- Paris CL (1989). The use of explicit user models in a generation system for tailoring answers to the user's level of expertise. In: Kobsa A, Wahlster W (eds.), *User Models in Dialog Systems*. Springer-Verlag, London, pp. 200–232.
- Peckham J (1993). A new generation of spoken dialogue systems: Results and lessons from the SUNDIAL project. *Proceedings of the 3rd European Conference on Speech Communication and Technology*, pp. 33–40.
- Pellom B, Ward W, Pradhan S (2000). The CU Communicator: An architecture for dialogue systems. *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP'2000)*, Beijing, China, November.
- Pellom B, Ward W, Hansen J, Hacioglu K, Zhang J, Yu X, Pradhan S (2001). University of Colorado dialog systems for travel and navigation. *Human Language Technology Conference (HLT-2001)*, San Diego, March.
- Phillips M (2003). User interface design for spoken dialogue systems. Invited talk. *Proceedings of the ISCA Workshop on Error Handling in Spoken Dialogue Systems*, p. 145.
- Pieraccini R, Dayanidhi K, Bloom J, Dahan JG, Phillips M (2003). A multimodal conversational interface for a concept vehicle. *Proceedings of the 8th International Conference on Speech Communication and Technology (Eurospeech2003)*, Geneva, September, pp. 2233–2236.
- Poesio M, Traum DR (1997). Conversational actions and discourse situations. *Computational Intelligence* 13(3).
- Poesio M, Traum DR (1998). Towards an axiomatization of dialogue acts. *Proceedings of Twendial'98, 13th Twente Workshop on Language Technology: Formal Semantics and Pragmatics of Dialogue*.
- Portele T (2003). Interaction modeling in the SmartKom system. *Proceedings of the ISCA Workshop on Error Handling in Spoken Dialogue Systems*, pp. 89–94.
- Portele T, Gorony S, Emele M, Kellner A, Torge S, te Vrugt J (2003). SmartKom-Home – An advanced multi-modal interface to home entertainment. *Proceedings of the 8th International Conference on Speech Communication and Technology (Eurospeech2003)*, Geneva, September, pp. 1897–1900.

- Pulman S (2002). Relating dialogue games to information state. *Speech Communication* 36:15–30.
- Raman TV (1997). *Auditory User Interfaces – Toward The Speaking Computer*. Kluwer Academic, Boston, MA.
- Raman TV, McCobb G, Hosn RA (2003). Versatile multimodal solutions: The anatomy of user interaction. *XML Journal* 4(4).
- Reilly R (1987) (ed.). *Communication Failure in Dialogue*. North-Holland, Amsterdam.
- Reiter E, Dale R (2000). *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge, UK.
- Richards MA, Underwood K (1984). Talking to machines. How are people naturally inclined to speak? In: Megaw ED (ed.), *Contemporary Ergonomics*. Taylor & Francis, London.
- Robertson S, Robertson J (1999). *Mastering the Requirements Process*. Addison-Wesley, London.
- Roe DG, Wilpon JG (1994) (eds.). *Voice Communication Between Humans and Machines*. National Academy Press, Washington, DC.
- Rooney E (2002). DKIT Voice Application. MSc Informatics Dissertation, University of Ulster.
- Roth SF, Mattis J, Mesnard X (1991). Graphics and natural language as components of automatic explanation. In: Sullivan J, Tyler S (eds.), *Intelligent User Interfaces*. ACM Press, New York, pp. 207–239.
- Rudnicky AI, Xu W (1999). An agenda-based dialog management architecture for spoken language systems. *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU99)*, pp. 3–7.
- Rudnicky AI, Thayer E, Constantinides P, Tchou C, Shern R, Lenzo K, Xu W, Oh A (1999). Creating natural dialogs in the Carnegie-Mellon Communicator system. *Proceedings of the 6th International Conference on Speech Communication and Technology (Eurospeech99)*, Budapest, Hungary, pp. 1531–1534.
- Rudnicky AI, Bennett C, Black A, Chotomongcol A, Lenzo K, Oh A, Singh R (2000). Task and domain specific modelling in the Carnegie-Mellon Communicator system. *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP2000)*, Beijing, China, October.
- Russell SJ, Norvig P (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ.
- Rutten P, Fackrell J (2003). The application of interactive speech unit selection in TTS systems. *Proceedings of the 8th International Conference on Speech Communication and Technology (Eurospeech2003)*, Geneva, September, pp. 285–288.
- Sacks H, Schegloff EA, Jefferson G (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50:696–735.
- Sadek D, de Mori R (1997). Dialogue systems. In: de Mori R (ed.), *Spoken Dialogs with Computers*. Academic Press, New York.
- San-Segundo R, Pellom B, Ward W, Pardo JM (2000). Confidence measures for dialogue management in the CU communicator system. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2000)*, June.

- Schank RC (1975). *Conceptual Information Processing*. North-Holland, Amsterdam.
- Scheffler K, Young S (2002). *Automatic Learning of Dialogue Strategy Using Dialogue Simulation and Reinforcement Learning*. HLT, San Diego, CA.
- Schegloff EA, Sacks H (1973). Opening up closings. *Semiotica* 8:289–327.
- Schiffrin D (1994). *Approaches to Discourse: Language as Social Interaction*. Blackwell, Oxford, UK.
- Schiffrin D, Tannen D, Hamilton H (eds.) (2001). *The Handbook of Discourse Analysis*. Blackwell, Oxford, UK.
- Searle JR (1969). *Speech Acts*. Cambridge University Press, Cambridge, UK.
- Searle JR (1975). Indirect speech acts. In: Cole P, Morgan JL (eds.), *Syntax and Semantics*, Vol. 3: *Speech Acts*. Academic Press, New York.
- Seneff S, Hurley E, Lau R, Pao C, Schmid P, Zue V (1998). Galaxy-II: A reference architecture for conversational system development. *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP'98)*, Sydney, Australia, November.
- Sensory Inc. (2002). Sensory's new speech technologies for interactive kiosk avatars make shopping easy. *Speech Technology Magazine NewsBlast*, March 12. Available at: <http://www.speechtechmag.com/pub/industry/159-1.html>
- Sidney S, Castellan NJ (1988). *Nonparametric Statistics for the Behavioural Sciences*. McGraw Hill, New York.
- Sinclair JM, Coulthard M (1975). *Towards an Analysis of Discourse*. Oxford University Press, Oxford, UK.
- Singh S, Litman DJ, Kearns MS, Walker MA (2002). Optimizing dialogue management with reinforcement learning: Experiments with the NJFun system. *Journal of Artificial Intelligence Research* 16:105–133.
- Sinha AK, Klemmer SR, Landay JA (2002). Embarking on spoken language NL interface design. *International Journal of Speech Technology* 5(2):159–169.
- Skantze G (2003a). Error detection in spoken dialogue systems. Available at: <http://www.speech.kth.se/~gabriel/courses/DialogueSystems5p/ErrorDetection.pdf>
- Skantze G (2003b). Exploring human error handling strategies: Implications for spoken dialogue systems. *Proceedings of the ISCA Workshop on Error Handling in Spoken Dialogue Systems*, pp. 71–76.
- Smith RW (1998). An evaluation of strategies for selectively verifying utterance meanings in spoken natural language dialog. *International Journal of Human-Computer Studies* 48:627–647.
- Smith RW, Hippi DR (1994). *Spoken Natural Language Dialog Systems*. Oxford University Press, New York.
- Stent AJ (2000). The Monroe Corpus, TR728 and TN99-2. Computer Science Department, University of Rochester, March.
- Stork DG (1998) (ed.). *Hal's Legacy: 2001's Computer as Dream and Reality*. MIT Press, Cambridge, MA.
- Stubbs M (1983). *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. Blackwell, Oxford, UK.

- Sturm J, den Os E, Boves L (1999). Dialogue management in the Dutch ARISE train timetable information system. *Proceedings of the 6th International Conference on Speech Communication and Technology (99)*, Budapest, Hungary, pp. 1419–1422.
- Suhm B (2003). Towards best practices for speech user interface design. *Proceedings of the 8th International Conference on Speech Communication and Technology (Eurospeech2003)*. Geneva, September, pp. 2217–2220.
- Sutton R, Barto AG (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.
- Swerts M, Litman DJ, Hirschberg J (2000). Corrections in spoken dialogue systems. *Proceedings of the 6th International Conference of Spoken Language Processing*, Beijing, China.
- Traum DR (1994). A Computational Model of Grounding in Natural Language Conversation. PhD thesis, University of Rochester.
- Traum DR (1996). A reactive-deliberative model of dialogue agency. In: Muller JP, Wooldridge MJ, Jennings NR (eds.), *Intelligent Agents III: Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96)*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Heidelberg.
- Traum DR (1999). Computational models of grounding in collaborative systems. *Proceedings of AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*. North Falmouth, MA. AAAI Press, Menlo Park, CA, pp. 124–131.
- Traum DR (2000). 20 questions for dialogue act taxonomies. *Journal of Semantics* 17(1):7–30.
- Traum DR, Allen JF (1994). Discourse obligations in dialogue processing. *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94)*, pp. 1–8.
- Traum DR, Larsson S (2001). The information state approach to dialogue modelling. In: *TRINDI Consortium, 2001, The TRINDI Book* (Draft). Available at: <http://www.ling.gu.se/projekt/trindi/publications.html>
- Traum DR, Larsson S (to appear). The information state approach to dialogue management. In: Smith R, Kuppevelt J (eds.), *Current and New Directions in Discourse and Dialogue*. Kluwer Academic, Dordrecht.
- TRINDI Consortium (2001). *The TRINDI Book* (Draft). Available at: <http://www.ling.gu.se/projekt/trindi/publications.html>
- Turing AM (1950). Computing machinery and intelligence. *Mind* 59:433–460.
- Turunen M, Hakulinen J (2001a). Agent-based error handling in spoken dialogue systems. *Proceedings of the 7th International Conference on Speech Communication and Technology (Eurospeech2001)*, Aalborg, Denmark, pp. 2189–2192.
- Turunen M, Hakulinen J (2001b). Agent-based adaptive interaction and dialogue management architecture for speech applications. Text, speech and dialogue. *Proceedings of the Fourth International Conference TSD*, pp. 357–364.



- VoiceGenie (2003). Available at: <http://www.voicegenie.com/pdf/Apps/VirtualAssistant.pdf>
- Wahlster W (1993). Verbmobil – Translation of face to face dialogs. *Proceedings of the Machine Translation Summit IV*, Kobe.
- Wahlster W, André E, Finker W, Profitlich HJ, Rist T (1993). Plan-based integration of natural language and graphics generation. *Artificial Intelligence Journal* 63:387–427.
- Wahlster W, Reithinger N, Blocher A (2001). SmartKom: Multimodal communication with a life-like character. *Proceedings of the 7th European Conference on Speech Communication and Technology*. Aalborg, Denmark, September, pp. 1547–1550.
- Walker MA (2000). An application of reinforcement learning to dialogue strategy selection in a spoken dialogue system for email. *Journal of Artificial Intelligence Research* 12:387–416.
- Walker MA, Litman DJ, Kamm C, Abella A (1997). PARADISE: A general framework for evaluating spoken dialogue agents. *Proceedings of the 35th Annual General Meeting of the Association for Computational Linguistics, ACL/EACL*, Madrid, Spain, pp. 271–280.
- Walker MA, Litman DJ, Kamm C, Abella A (1998). Evaluating spoken dialogue agents with PARADISE: Two case studies. *Computer Speech and Language* 12(3):317–347.
- Walker MA, Wright J, Langkilde I (2000a). Using natural language processing and discourse features to identify understanding errors in a spoken dialogue system. *Proceedings of the 17th International Conference on Machine Learning*.
- Walker MA, Langkilde I, Wright J, Gorin A, Litman DJ (2000b). Learning to predict problematic situations in a spoken dialogue system: Experiments with How may I help you? North American Meeting of the Association of Computational Linguistics.
- Walker MA, Rudnicky AI, Prasad R, Aberdeen J, Owen Bratt E, Garofolo J, Hastie H, Le A, Pellom B, Potamianos A, Passonneau R, Roukos S, Sanders G, Seneff S, Stallard D (2002). DARPA Communicator: Cross-system results for the 2001 evaluation. *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP2002)*, Denver, CO, pp. 269–272.
- Ward W (1991). Understanding spontaneous speech: The Phoenix system. *Proceedings of the ICASSP'91*, pp. 365–367.
- Webber BL (2001). Computational perspectives on discourse and dialogue. In: Schiffrin D, Tannen D, Hamilton H (eds.), *The Handbook of Discourse Analysis*. Blackwell, Oxford, UK.
- Weinschenk S, Barker DT (2000). *Designing Effective Speech Interfaces*. Wiley, New York.
- Weizenbaum J (1966). ELIZA – A computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery* 9:36–45.
- Whittaker SJ, Attwater DJ (1996). The design of complex telephony applications using large vocabulary speech technology. *Proceedings of the 4th Interna-*

- tional Conference on Spoken Language Processing (ICSLP'96)*, Philadelphia, PA, pp. 705–708.
- Whittaker S, Walker MA, Maloor P (2003). Should I tell all?: An experiment on conciseness in spoken dialogue. *Proceedings of the 8th International Conference on Speech Communication and Technology (Eurospeech2003)*, Geneva, September, pp. 1685–1688.
- Wilensky R (1981). PAM. In: Schank RC, Riesbeck CK (eds.), *Inside Computer Understanding*. Lawrence Erlbaum, Hillsdale, NJ.
- Winograd T (1972). *Understanding Natural Language*. Academic Press, New York.
- Winther M (2001). Leveraging Web Technologies to Achieve Customer-Centric Contact Centers. Voice Web Today. Available at: <http://www.telera.com/news/newsletter/viewpoint.html>
- Woods WA, Nash-Webber BL, Kaplan RM (1972). The Lunar sciences natural language system final report. BBN Report 3438. Bolt, Beranek and Newman, Cambridge, MA.
- Wooffitt RC (1991). The moral and inferential properties of silence: A warning note to designers of intelligent speech systems. ESPRIT Project P2218 (SUNDIAL) Working Paper.
- Wright Hastie H, Poesio M, Isard S (2002). Automatically predicting dialogue structure using prosodic features. *Speech Communication* 36:63–79.
- Yankelovich N, Levow G, Marx M (1995). Designing SpeechActs: Issues in speech user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Denver, CO, pp. 369–376.
- Young SJ (2000). Probabilistic methods in spoken dialogue systems. *Philosophical Transactions of the Royal Society (Series A)* 358(1769):1389–1402.
- Young SJ (2002). Talking to machines (Statistically speaking). *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP2002)*, Denver, CO.
- Zue V (1997). Conversational interfaces: Advances and challenges. *Proceedings of the 5th International Conference on Speech Communication and Technology (Eurospeech97)*, Rhodes, Greece.
- Zue V (1999). Talking with your computer. *Scientific American*, August 19, 1999.

# Index

---

## A

- ABNF (Augmented Backus-Naur Form), 391–392
- Acoustic confidence score, 87, 97–98, 354–356, 360, 361, 371, 374, 375; *see also* recognition confidence
- Acoustic model, 84–86, 88–89
- Adaptable dialogue systems, 362, 363–364; *see also* adaptive dialogue systems
- Adaptive dialogue systems, 4, 347, 349, 354, 362–371, 375, 376–377, 385
- Adjacency pair, 65–67, 70–71
- Alphadigit recognition, 184–187, 253–254
- Ambiguity, 91, 103, 118, 122
- Anaphoric reference, 29, 46–47, 48, 123, 330, 352; *see also* reference (discourse)
- Animated agents, 8, 18, 22, 171, 225, 229, 323, 328, 383
- Apache Tomcat, 279, 288, 296, 397, 399
- Apple, 37, 401
- Artificial intelligence, 1, 2, 17, 24–25, 27–28, 30, 31, 34, 35, 36, 41, 46, 116–117, 349, 375
- ARPA, 35, 36, 156
- ASP (Active Server Pages), 232, 278, 316
- ASR confidence; *see* acoustic confidence score
- ATIS, 36, 37, 56, 109, 131, 339
- AT&T, 37, 38, 358, 371, 404
- AutoTutor, 23

## B

- Baldi, 22, 191, 206, 225, 230
- Barge-in, 54, 90–91, 133, 143, 255–257, 277
- Bayesian networks, 341, 343–345
- Bayesian receptionist, 341, 345–347; *see also* Conversational Architectures Project

- BBN, 324
- BDI (belief, desire, intention), 30, 51, 350
- BeVocal, 37, 241, 242, 391, 393–394, 407

## C

- Centre for Speech Technology (Edinburgh), 37, 324
- Chatterbot, 24, 42–43; *see also* simulated conversation
- Circuit-Fix-It-Shop, 21–22, 370–371
- Cisco, 38, 304
- Clarification, 9, 55, 60, 68, 71, 108, 109, 326, 338, 340, 344, 345, 372; *see also* clarification subdialogue; grounding; repair
- Clarification query; *see* clarification
- Clarification request; *see* clarification
- Clarification subdialogue, 108, 118, 340
- CMU Communicator, 324, 330–331, 347, 348, 403
- CMU Pronouncing Dictionary, 84, 194, 319
- Coherence, 14, 123
- Collaborative activity, dialogue as, 52–55, 75, 341
- Command and control, 8, 10–11
- Command, global, 233, 257–260, 317
- Command, navigational, 135, 139, 145, 161, 233, 257–260
- Common ground, 350–353; *see also* grounding
- Common ground unit (CGU), 71–72; *see also* grounding
- Computational linguistics, 2, 93, 126, 349
- Comverse, 38, 304
- Concept accuracy (CA), 155–156, 356–357, 358
- Concept density, 156
- Conceptual dependency theory, 27, 95

- Confidence; *see* recognition confidence  
Confirmation; *see* verification  
Consistency, 102, 139, 145, 146, 361  
Context, 14, 27–28, 29, 49–51, 55, 59, 60,  
73, 87, 93, 101, 117, 125, 158, 326,  
333, 370–371, 381–382, 384, 385  
dialogue, 46, 116, 329–330, 349–350,  
358, 366  
discourse, 14, 75, 335, 337–339, 350, 358  
linguistic, 82  
Context shift, 358, 361  
Context-free grammar, 92–94, 316, 325  
Contextual appropriateness, 157–158  
Conversation, 8, 12, 36, 41, 45–46, 51, 53,  
61, 66, 67, 70, 71, 108, 132, 341, 342;  
*see also* dialogue  
Conversation analysis (CA), 58, 65–66,  
70–71, 76  
Conversational agent, 23, 126, 323,  
336–337, 380  
Conversational Architectures Project  
(Microsoft), 323, 341–347, 348, 403  
Conversational competence, 124; *see also*  
discourse obligations  
Conversational game theory, 64–65, 68, 72  
Conversational implicature, 51–52  
Conversational Interaction and Spoken  
Dialogue Research Group  
(Rochester), 332–333, 347, 402  
Conversational interface, 22, 36, 126, 384,  
385; *see also* spoken dialogue  
system  
Conversational maxims, 51–52  
Conversational move, 60, 64–65, 67, 68,  
69, 351  
Cooperative principle, 51–52  
Corpora, 36, 61, 69, 82, 126, 131, 160, 406  
Correction rate, 157  
CSELT (Italy), 37  
CSLR (Center for Spoken Language  
Research), 22, 37, 201, 206–208,  
229, 230, 323, 324, 325, 326, 328,  
402; *see also* CSLR Reading Tutor;  
CU Animate; CU Communicator  
CSLR Reading Tutor, 229, 230  
CSLR Sonic Speech Recognizer, 325  
CSLU (Center for Spoken Language  
Understanding), 22, 37, 146, 402;  
*see also* CSLU toolkit, CSLU  
Vocabulary Tutor  
CSLU toolkit, 3, 22, 98, 111, 112, 118, 121,  
145–148, 163–203, 205–230, 255,  
395, 406  
CSLU Vocabulary Tutor, 22, 229  
CU Animate, 22, 206–208, 212, 230  
CU Communicator, 323, 324–330, 332,  
347, 348, 403, 406  
CU Conversational Agent Toolkit,  
328–329  
Cues, 359–360, 361  
prosodic, 53, 356, 387  
visual, 345–346
- ## D
- Daimler-Chrysler research, 324  
DAMSL, 61–64, 65  
SWBD-DAMSL, 63–64, 69  
DARPA Communicator, 14, 36–37, 98,  
323–324, 327–328, 333, 334, 382;  
*see also* CMU Communicator; CU  
Communicator; Galaxy  
architecture; Queen's  
Communicator  
DARPA, spoken language program, 36  
Database access, 28–29, 121–123, 135,  
139, 148, 197–201, 278–288, 326,  
372  
Database query systems, 28–29  
Data entry, 11–12  
Decision making, 341–345  
Department of Language and Speech  
(Nijmegen), 37, 324, 402  
Design, 139–146  
Development lifecycle; *see* dialogue  
engineering  
DFKI (Germany), 37, 381  
Dictation, 12–13, 87  
Dialogue, 2, 3, 9, 12, 13–15, 45–77; *see also*  
conversation  
definition, 45–46  
as extended discourse, 46–49, 75  
as purposeful activity, 49–52, 75  
as rational action, 72–73  
as social behaviour, 70–71  
Dialogue act, 60–65  
backward-looking, 61, 62–63, 352–353  
forward-looking, 61–62, 352  
Dialogue applications  
educational, 3, 22–23  
entertainment, 23–24  
games, 23–24  
information retrieval, 12, 19–20  
intelligent tutoring, 22–23  
problem solving, 16, 20–22, 41, 107,  
116–117, 330–331, 332–339  
Dialogue coding, 59–69  
Dialogue control, 14, 80–81, 107–108,  
111–117, 125  
agent-based, 116–117, 124  
finite-state based, 111–112, 116, 124

- frame-based, 30–31, 113–116, 124, 326, 327, 328
  - Dialogue engineering, 3, 129–161
  - Dialogue grammar, 69–70, 107
  - Dialogue history, 123, 332, 335, 343, 358, 365–366, 371
  - Dialogue management, 3, 9, 16, 79, 107–126, 326, 328–332, 358, 372; *see also* Dialogue control
    - knowledge sources, 123–124
    - issues for developers, 124–126
    - object-oriented, 331–332
    - task record, 124
  - Dialogue move; *see* conversational move
  - Dialogue obligations; *see* discourse obligations
  - Dialogue policy, 372, 374–375; *see also* reinforcement learning
  - Dialogue processes, 69–75
  - Dialogue projects and systems
    - Annie, 354–356
    - ARISE, 36
    - BASEBALL, 25
    - Circuit-Fix-It-Shop, 21–22
    - CONVERSE, 34
    - Danish Dialogue Project, 132, 136
    - D’Homme, 9, 109
    - DIPPER, 376, 403
    - DISC, 36, 136, 403
    - EAGLES, 136, 156
    - ELIZA, 32–33
    - Finnish Interact project, 365–367
    - GUS, 30–31
    - How May I Help You (HMIHY), 358, 360–361, 404
    - Jupiter, 12, 404
    - Kyoto City Bus Information, 367–369
    - LUNAR, 28
    - NICE, 23–24
    - NJFun, 371–375
    - PAM, 28
    - PARRY, 33–34
    - Pegasus, 12
    - Philips Automatic Train Timetable Information System, 20, 36, 115
    - POLITICS, 28
    - RAILTEL, 36
    - QUALM, 28
    - SAM, 27–28
    - SHRDLU, 26–27
    - SpeechActs, 131
    - STUDENT, 25
    - SUNDIAL, 36
    - TOOT, 356, 364–365
    - Verbmobil, 36, 404
    - Voyager, 12, 403
  - Dialogue state, 371–374; *see also* reinforcement learning
  - Dialogue strategy, 157–158, 347, 349, 355–356, 357, 362–365, 385
    - optimal strategy, 371–375
  - Dialogue structure, 59–69, 76
  - Digit recognition, 11, 135, 179–181, 182, 199, 202, 247–249
  - Directed dialogue, 4, 107, 108–109, 114–115, 134, 145, 163, 231–261, 269, 384
  - Discourse, 9, 13–14, 25, 27–28, 29, 46–49, 67, 76
  - Discourse analysis, 60, 76
  - Discourse history; *see* dialogue history
  - Discourse obligations, 124, 335, 336–337, 350–353
  - Discourse Representation Structures, 351
  - Discourse Representation Theory, 351
  - Discourse segment, 65, 67–68, 75
  - Discourse unit (DU), 352–353
  - Disfluency, 56–57, 358
  - Dispreferred response, 66–67, 360; *see also* adjacency pair
  - Document type definition (DTD), 235, 394
  - DTMF, 107, 133, 135, 181–184, 232, 249–253, 278, 291, 398
  - Dynamic programming, 85–86
- E**
- Echo cancellation, 10
  - ECMAScript, 233, 260–261, 278
  - E-form, 114, 115–116; *see also* dialogue control, frame-based
  - Ellipsis, 29, 101, 123, 330
  - ELNET, 18
  - Email reader, 12, 13, 160
  - Embedded appliances, 8–10, 16, 385–386
  - Embodied conversational agents; *see* animated agents
  - Error handling, 4, 13, 14–15, 117, 125, 349, 375, 376, 386–388
    - detection, 354–360
    - prediction, 360–361
    - prevention, 354
    - recovery, 354, 361–362
  - Evaluation, 154–160, 327–328, 340–341, 369
    - metrics, 156–157
    - spoken dialogue systems, 156–160
    - system performance, 155–160
    - user acceptance, 154–155

- Exchange (dialogue), 65, 67, 72
- External communication, 80, 121–123, 125–126; *see also* database access  
ambiguous and indeterminate data, 122–123  
vocabulary, 122
- F**
- Festival, 304
- Fluent Animated Speech; *see* Sensory, Inc.
- Form interpretation algorithm (FIA); *see* VoiceXML, 3, 114–115, 237
- Frames, 30–31, 113–116, 124, 325–326, 327, 329, 347
- Functional specification; *see* requirements specification
- G**
- Galaxy architecture, 12, 324–326, 404
- Game Commander, 23
- Grammar, 11, 144–145; *see also* Recognition grammar  
DTMF grammar, 145  
finite-state, 86–87
- Graphical user interface (GUI), 133, 141, 231, 319, 333, 335  
and speech interface, 129, 130, 131–132
- Grounding, 54–55, 71–72, 108, 117–121, 341, 344–345, 351–353; *see also* Verification
- GSL (Grammar Specification Language), 391
- H**
- HAL (Home Automated Living), 8
- Hands-busy/eyes-busy, 10, 379, 384
- HCRC Map Task, 402  
coding scheme, 64–65  
corpus, 131
- Help (system), 145–146, 362, 365, 384
- Hey Anita, 37, 407
- Hey You, Pikachu, 23
- Hidden Markov Model, 85–86
- Hill-Rom Enhancemate, 10
- I**
- IBM, 2, 9, 18, 37, 38, 292, 324
- IBM Multimodal Toolkit, 150, 292, 294–303, 319, 400, 406
- IBM Reusable Dialog Components, 253, 295, 398
- IBM WebSphere Voice Server SDK, 148–149, 160, 235, 241, 253, 391, 393, 397–398, 406
- IBM WebSphere Voice Toolkit, 148–149, 235, 253, 398–399  
Content Assist tool, 148–149, 294  
Grammar Development Tool, 294  
Grammar Test Tool, 294  
Pronunciation Builder Tool, 295
- Ill-formed input, 29, 70–71, 91, 94, 96–97, 118, 340
- Imitation game, 31–32
- Implementation, 146–150
- Inference, 27–28, 30, 51–52, 58–59, 66, 70, 73, 341, 343, 346
- Information state theory, 4, 55, 73–75, 335, 347–348, 349–353, 371, 375–376
- Initiative (dialogue), 9, 107, 108–111, 125; *see also* Directed dialogue; Mixed initiative dialogue; User initiative
- Intel, 2, 37, 38, 304
- Intention, 30, 35, 41, 51, 61, 65, 68, 73, 74, 124, 332, 342, 353
- Interactive kiosk, 8
- Interactive voice response (IVR), 40, 231
- Internet  
and speech, 1, 8, 13, 17, 18
- J**
- Jaspis, 365–366
- JavaScript; *see* ECMAScript
- JSGF (Java Speech Grammar Format), 235, 271, 299, 391, 394
- JSML (Java Speech Markup Language), 255
- JSP (Java Server Pages), 232, 278, 279–288, 399
- K**
- KQML (Knowledge Query and Manipulation Language), 334
- KTH (Sweden), 37, 324, 402
- L**
- Language generation, 3, 80, 99–102, 326, 337, 365–367  
document planning, 99–100  
issues for developers, 102  
microplanning, 100–101  
surface realisation, 101
- Language model, 87–88, 105
- LIMSI (Paris), 37, 402
- Linguistic Data Consortium, 36
- Loebner Prize, 31–32, 34
- Lombard effect, 54
- Lucent, 38

**M**

- Machine learning, 4, 98–99, 347, 349, 356–358, 360–361, 364, 368, 385
- Markov Decision Process (MDP), 372, 375
- Mental state, 51, 73, 74, 350–351
- Microsoft, 2, 37, 38, 304, 324, 341, 345, 401, 406
- Microsoft (MS) Access, 28, 197, 198, 395
- Microsoft .NET Speech SDK, 150, 304–307, 316–319, 400
  - Grammar Editor, 316–317
  - Internet Explorer Speech Add-In, 309–315
  - Prompt Editor, 317–318
  - QA controls, 316, 317
  - Speech Control Editor, 317
  - Speech Debugging Console, 318
- Microsoft Visual Studio .NET, 150, 304, 309, 316–318, 400
- Miscommunication, 13, 14–15, 54, 341, 354–356, 369–370
- Misrecognition, 15, 41, 55, 143, 152, 156, 159, 228, 308, 356, 364
- Misunderstanding; *see* miscommunication
- MIT, 12, 14, 26, 324 (or *see* spoken languages group); *see also* Spoken Language Systems Group
- MITRE, 24, 326
- Mixed-initiative dialogue, 4, 107, 108, 110–111, 114–115, 117, 134, 145, 263–266, 269–270, 303, 323, 333, 362, 363, 371, 373–375
- Motorola, 37, 38, 292
- Move; *see* Dialogue move
- Multimodal development platforms, 150
- Multimodal dialogue, 3, 4, 205–228, 291–319, 341, 379–385
  - error handling, 381–382
  - fusion, 382
  - input, 379, 380
  - modality integration, 382
  - output, 379, 380–381
  - in vehicles, 384–385
- Multimodal dialogue systems
  - COMET, 381
  - CUBRICON, 380
  - MATCH, 380, 382
  - PPP, 381
  - Quick-Set, 380, 382
  - SAGE, 381
  - SMARTKOM, 383–384, 404
  - WIP, 381

**N**

- Natural Interactive Systems Laboratory (Denmark), 37, 324, 402
- Natural language processing, 1, 2, 18, 105, 117, 358
- Natural language understanding, 3, 9, 79, 91–99, 120–121, 325–326, 327, 333, 346, 361
  - issues for developers, 98–99
  - syntactic approaches, 92–94, 105
- Navigation; *see* Command, navigational
- N-gram; *see* Language model
- NIST (National Institute of Standards), 328
- Noise, 10, 81, 83, 379, 385, 386
- Non-verbal behaviour, 58, 383
- Nuance Communications, 9, 18, 37, 39, 40, 253, 391, 401, 406

**O**

- ODBC, 197–198, 210, 280, 395–396
- OpenSALT, 304
- OpenSpeech DialogModules, 253
- Opera, 292, 299, 319

**P**

- PARADISE, 158–160, 372
  - Parsing, 87, 96–97, 103, 158, 339, 340, 382
  - Parser confidence, 358, 370
  - PDA, 8, 38, 40, 291, 292, 318, 386
    - speech-enabled, 13, 150
  - Perceptual Science Laboratory, 22
  - Personal data assistant; *see* PDA
  - Philips, 17, 38, 304, 324
  - Phoenix, 97, 325–326
  - Phoneme, 82, 83–85, 89–90, 102, 104, 195–196, 380
  - Planning, 30, 72–73, 332–333, 381
  - Plan recognition, 51, 72–73, 336–337
  - Pragmatics, 76
  - Preference; *see* adjacency pair
  - Prompt design, 125, 133–134, 143–144, 166, 167, 349, 354, 387
    - incremental prompts, 143–144
    - leading prompts, 144, 362
    - tapered prompts, 143
  - Prosody, 104, 356–358, 360, 385, 387
- Q**
- Quartet, 341–345; *see also* Conversational Architectures Project
  - Queen's Communicator, 324, 331–332, 347
  - Query density, 156

**R**

RAD (Rapid Application Developer), 3, 147–148, 163–202, 205–230, 233, 279

- alphadigit recognition, 184–187
- animated face, 171, 206–208, 212–216
- background colours, 225–227
- canvas, 164
- database connection using TclODBC, 197–201
- data capture and playback, 228–229
- default repair subdialogue, 174–176
- dialogue graph, 165, 172–174, 175, 176–177, 179–180, 182–184, 185, 199–201, 210–211, 213–216, 218, 221, 223
- dictionaries, 227
- digit recognition, 179–181
- DTMF, 181–184
- dynamic recognition, 196–197
- emotions, 211–216
- garbage, 175–176, 188
- Listbuilder, 217–220
- login dialogue, 208–211
- media object, 220–223
- out-of-vocabulary rejection, 227–228
- prompts, 166, 167
- pronunciation, 168, 192, 194–196
- recognition grammar, 187–191
- recognition problems, 166
- recognition results, 174–176
- recognition vocabulary, 167
- recording speech, 225
- speech output, 191–192
- states, 168–169
- subdialogues, 176–179, 214–216
- trailing silence duration, 180–181, 188
- SayAs, 191–192
- TTS, 166, 171, 186, 191, 208–209, 219, 224–225
- verification, 169–170, 171

Reasoning, 15–16, 116–117, 332–333

Recognition confidence, 354–356, 357, 364–365, 373, 384; *see also* Acoustic confidence score

Recognition errors, 13, 15, 82, 386–388; *see also* Misrecognition

Recognition grammar, 3, 89, 120–121, 187–191, 270–274, 354, 357

Reference (discourse), 14, 31, 100–101

Reinforcement learning, 349, 371–375, 376, 377; *see also* Machine learning

Repair (conversational), 54–55, 57, 71, 73, 344–345; *see also* Clarification

Requirements analysis, 129–135

Requirements specification, 135–138

Response generation, 80, 102, 376

Rhetorical structure theory, 100

RIPPER, 357–358

Rochester systems; *see* Conversational Interaction and Spoken Dialogue Research Group

**S**

SABLE, 191, 255

SALT, 1, 38–39, 291, 304–319, 379, 385, 400

- and VoiceXML, 307
- elements,
  - <dtmf>, 307
  - <listen>, 307, 308, 312, 313–314, 315
  - <prompt>, 307–308, 311, 313
  - <smex>, 307
- event wiring, 308–309
- namespace declarations, 311
- script, 314–315
- XML grammar, 312, 314

SALT Forum, 38–39, 304, 405

Scansoft, 9, 18, 23, 37, 253, 388, 401

Scripts, (in artificial intelligence), 27–28, 34, 35

Semantic grammar, 94–96

Sensory Inc., 8, 11, 206, 208, 209

Sentence accuracy (SA), 155

Sentence understanding rate (SU), 155, 156

Signal processing, 83–84

Silence (in conversation), 58–59

Simulated conversation, 31–35

SISR (Semantic Interpretation for Speech Recognition), 319

SML (Semantic Markup Language), 316

Space Odyssey 2001, 1, 16, 17

Speaker verification, 8

Speech acts, 50–51, 60, 76, 326, 335, 340, 352

SpeechMania, 115

SpeechObjects, 253

Speech recognition, 1, 3, 8, 10, 14, 16, 79, 81–91, 105, 324–325, 328, 344, 346, 363, 364, 385

- enrolment, 87–88
- integration with language understanding, 97–98
- issues for developers, 88–91
- out-of-grammar, 386–388
- performance, 105, 130, 138, 151, 155, 228, 270, 277, 325, 339–340, 356, 357, 360, 362, 364, 379
- speaking mode, 88



- tuning speech recognition, 88–89
  - variability, 82–83
  - Speech technology, 2, 9, 10, 13, 16, 17, 35–36
    - benefits, 39–40
    - challenges, 40–41
    - market, 38
    - and the telephone, 1, 13, 16, 17
  - Speech Technology Magazine, 17
  - SpeechWorks, 38, 304, 406
  - Sphinx (CMU), 37, 304, 324, 339–340
  - Spoken Dialogue Research Group (Rochester), 16, 37
  - Spoken dialogue system, 1, 2, 3, 14, 17, 18, 79–81, 126, 341
  - Spoken dialogue technology, 1, 2–4, 16, 17, 126, 349, 375, 377, 379, 385–388
    - commercial deployment, 37–41, 42, 45, 386–388
  - Spoken language requirements, 129, 131–133
  - Spoken Language Systems Group (MIT), 12, 37, 126
  - SRG XML (Speech Recognition Grammar), 391
  - SRI, 37, 324
  - SSML (Speech Synthesis Markup Language), 255
  - Star Trek, 7, 16
  - State transition network; *see* Dialogue control, finite-state based
  - Subdialogue, 9, 176–179, 214–216, 243–247
  - Suede design tool, 160, 406
  - SUN Microsystems Laboratories, 131, 132
  - System-directed dialogue; *see* System-initiative
  - System in the loop method, 132–133
  - System-initiative, 305, 362, 363, 366, 368, 371, 373–375, 387; *see also* Directed dialogue
- T**
- Talking head, 3; *see also* Animated agents, Baldi
  - Task hierarchy diagram (THD), 139–142
  - Tcl, 170, 192–194, 202, 205, 223–224
  - TclODBC, 197–199, 202, 279, 395
  - Tellme, 37, 407
  - Testing, 150–154
    - black box testing, 151
    - integration testing, 152–153
    - system testing, 154
    - unit testing, 151–152
    - validation testing, 153–154
    - white box testing, 150–151
  - Text-to-speech synthesis, 3, 8, 80, 102–104, 254–257, 326; *see also* RAD TTS, VoiceXML TTS
    - concatenative speech synthesis, 104
    - issues for developers, 104
    - tagging, 103
  - Timing (in dialogue), 387
  - Topic shift, 47–49, 331
  - Topic tracking, 14, 47–49
  - TRAINS, 131, 323, 332–333, 339–340, 404
  - Transaction, 68–69
  - Transaction success, 154, 157–159, 373
  - Transaction time, 157
  - TRINDI, 350, 376, 404
  - TRINDIKIT, 350, 376, 406
  - TRIPS, 16, 323, 332–341, 404
  - TTS; *see* text-to-speech synthesis
  - Tucker-Maxon Oral School, 22, 201, 205, 228
  - Turn-taking (conversational), 53–54, 59, 75, 124, 338
- U**
- Uncertainty, 333, 341–345
  - Unisys, 37, 406
  - Update rules, 351, 352–353
  - Use case analysis, 129–131
  - User initiative, 107, 108, 109–110, 363, 364, 366, 373–375
  - User model, 124, 133, 365–366, 367–369
  - User profile; *see* User model
  - Utterances (in dialogue), 13–14, 55–59
    - form, 56–57
    - prosody, 57–58
    - utterance relations, 13–14
- V**
- Verification, 118–121, 125, 135, 146, 169–171, 331–332, 349, 354–355, 359–360, 361, 363–365, 369–371, 373–375, 384–385
    - explicit, 119–120, 146, 332, 362, 363, 364, 366, 373–375, 387
    - implicit, 120–121, 331–332, 359, 363, 364, 366
    - in RAD, 169–170, 171
    - in VoiceXML, 239–241
  - Vialto, 13
  - Virtual personal assistant, 13
  - Viseme, 380
  - Vocabulary size, 11, 12, 88
  - Voice activation, 9, 11, 13
  - Voice-Enabled Virtual Assistant, 13

- VoiceGenie, 18, 37, 407
  - Voice Web, 13, 17, 38–39, 42
  - VoiceXML, 1, 2, 4, 37, 38, 95, 98, 110, 114–115, 118, 161, 231–261, 263–288, 305, 316, 362, 385
    - alphanumeric recognition, 253–254
    - application, 235, 236
    - ASR (automatic speech recognition), 232
    - audio prompts, 257
    - barge-in, 255–257
    - conditional (cond), 240–241, 246–247, 260
    - dialog, 236
    - digit recognition, 247–249
    - document, 235, 236
    - DTMF, 249–253, 278
    - dynamic, 4, 278–288
    - form interpretation algorithm (FIA), 3, 114–115, 237, 266–270
    - grammar, 4, 264–265, 270–274
      - built-in,
        - boolean, 240
        - currency, 252
        - date, 252
        - digit, 247–249
        - number, 252
        - phone, 252
        - time, 252
      - dynamic, 284–288
      - formats, 391–394
      - form-level, 263, 264, 393–394
      - scope, 276–277
      - semantic tag, 272–273, 392
    - language, 233
    - mixed-initiative, 263–266, 269–270, 393
    - navigation, 257–260
    - property, 277–278
      - bargein, 277
      - caching, 278
      - confidence, 277
      - timeout, 278
    - repair, 241–243; *see also* VoiceXML tags, <noinput>, <nomatch>
    - root document, 259–260
    - session, 236
    - speech markup, 255–257
    - speech output, 254–257
    - TTS, 232, 235, 252, 254–255
    - variable, 240, 274–276
      - scope, 274–276
    - verification, 239–241
  - VoiceXML browser, 232–233, 397
  - VoiceXML Forum, 38, 405
  - VoiceXML platforms, 148–150
    - standalone, 148–149
    - web-based, 149–150
  - VoiceXML server; *see* VoiceXML browser
  - VoiceXML specification, 252, 253, 391, 397
  - VoiceXML tags
    - <block>, 239, 275
    - <clear>, 240, 268
    - <dtmf>, 249–253
    - <field>, 235
    - <filled>, 237, 238, 240–241, 264, 267, 275, 280
    - <form>, 235, 264, 275
    - <goto>, 258, 259
    - <grammar>, 235, 271, 276
    - <initial>, 263–265, 269, 270
    - <link>, 259
    - <menu>, 238–239
    - <noinput>, 241, 242–243, 362
    - <nomatch>, 241–243, 362
    - <param>, 245–246
    - <prompt>, 234, 251, 255, 257
    - <submit>, 259
    - <subdialog>, 243–247, 258
    - <var>, 245–247, 274–276
    - <vxml>, 234
  - Voice Web Solutions, 304, 401
  - Volere, 135, 136
  - Voxeo, 37, 407
  - Voxpilot, 37, 391, 407
- ## W
- Web server applications, 278–288
  - Wizard of Oz (WOZ) method, 132–133, 160, 361–362
  - Word accuracy (WA), 155, 156, 328, 356–357
  - Word error rate; *see* Word accuracy
  - W3C (World Wide Web Consortium), 2, 38, 253, 255, 292, 304, 391, 405
- ## X
- XHTML + Voice, 4, 150, 291–303, 318–319, 379, 400
    - header and namespace declarations, 294
    - scripts, 300
    - visual component, 295–296
    - voice component, 296–298
  - XML, 101, 150, 233, 235, 278, 293, 316, 382, 391–394
  - Xybernat, 9