What we can learn from Dialogue Systems that don't work
On Dialogue Systems as Cognitive Models

David Schlangen
Department of Linguistics
University of Potsdam, Germany
das@ling.uni-potsdam.de

Abstract

In the 'real world', dialogue systems typically are made to work long days in call-centres of airlines and banks, fielding customer queries (and often inviting customer rage). In academia, a strong line of research is aimed at making such systems better at such tasks (in the hope of reducing customer annoyance). Here, I want to explore potential uses of spoken dialogue systems not as members of the workforce but in the lab, as a tool for the cognitive sciences. I argue that dialogue systems can be employed as situated, implemented computational models of language-capable agents; models whose predictions can be evaluated in real-time in ecologically valid settings, by human conversant. I sketch a methodology for building such models, propose areas where they can best be employed, and discuss the relations between research in this direction and more applied research.

1 Introduction

(Pieraccini and Huerta, 2005) recently noted that "there are three different lines of research in the field of spoken dialogue", one focusing on "understanding human communication, the second on designing the interface for usable machines, and the third on building those usable machines". Collapsing the latter two classes into one, we may label these views the *tool-for-understanding* and the *getting-things-done* approaches.[1]

Interestingly, (Pieraccini and Huerta, 2005) don't give any references for whom they see as representing the first line of research. And on closer inspection of the literature, there indeed seems to be little work in the dialogue systems community that would identify itself as belonging solely to the *tools-for-understanding* camp (it's a different matter in the embodied agents community).[2] In this paper, I'd like to explore the problems and potential of the *tool-for-understanding* direction and its relation to the *getting-things-done* camp.

The paper is structured as follows: First, I briefly review what computational cognitive models are and discuss how dialogue systems can be seen as a special class thereof. Then, I discuss a methodology for employing SDSs to address cognitive questions, and areas that seem particularly amenable to this methodology, given the current state of the technology. I then discuss a number of possible objections against the proposed use of dialogue systems. I close with some thoughts on the relation between the different uses for dialogue systems, and a general discussion.

2 Dialogue Systems as Cognitive Models

How can dialogue systems, with all their well-known technical problems and clumsy dialogue behaviour possibly function as models of cognitive abilities, and of which ones in any case? Before I address these questions, let us backtrack a bit and briefly review what cognitive models actually are.

2.1 Levels of Analysis in Cognitive Models

In the most abstract sense, a model in the cognitive sciences can be seen as a function from an agent's inputs to its outputs—typically, but not necessar-

[1] This of course reflects a classic dichotomy within the field of artificial intelligence which goes by many names: engineering vs. "empirical science concerned with the computational modeling of human intelligence" (Jordan and Russell, 1999); or, wrt. dialogue systems, "simulation" vs. "interface" (Larsson, 2005), or just simply applied vs. pure research.

[2] Recent examples of systems that seem to fall more on the *tool-for-understanding* side (but that do not make clear whether they see themselves as such) are (Allen et al., 1995; Allen et al., 2000; DeVault and Stone, 2009; Skantze and Schlangen, 2009).

Proceedings of the 13th Workshop on the Semantics and Pragmatics of Dialogue, June 24–26, 2009, Stockholm, Sweden.

ily, percepts and behaviours, respectively. In non-trivial cases, this function will depend in some way to the input (i.e., not be constant), and so can be seen as specifying an *information processor*.

As Marr (1982) pointed out in his seminal work on vision, such a function can be specified in different ways, which address different analytical interests; his classification is shown here in Table 1. A computational model is one which focuses on the problem that is being solved by the processor, i.e. only on the function in a mathematical sense. A representational model adds concerns about the exact way the processor computes the function; an implementational model also worries about the physical details of the processor.

A popular and fruitful recent line of research puts a further constraint on models on the computational level. With the, often tacit, assumption that natural behaviours have evolved to be near-optimal, they assume that agents act *rational*, i.e. that they solve their computational problems in an optimal way (minimizing their cost, maximizing their gain), given the available information (Anderson (1991), see also Chater and Oaksford (2008) for a recent overview). This direction has the advantage of offering a clear mathematical basis for computational modeling (probability theory, and more specifically Bayesian belief updating); we will discuss below to what extent it can support dialogue modeling.

Because it offers a convenient vocabulary to talk about inputs, outputs, and everything in between, we introduce here some central notions. The task of the agent can in such a model be stated clearly: it is to find that action a_t, given the observations of the world o_{t-1}, that has the best chance of bringing the world to a desired state s_{t+1}.

I now try to situate dialogue systems within this view of cognitive modelling.

2.2 Dialogue Systems: Situated Computational Whole-Agent Models

First, a few words on what I mean by "dialogue system". Often, the term is used specifically for mono-modal, *voice-only* systems that do rather limited practical tasks, and is used somewhat in opposition to *conversational agent* (seen as more capable, but less oriented towards practical applications), *multi-modal system* (with more modalities available to it) or *embodied conversational agents* (with a simulated or real "body", and consequently also more modalities). I do not intend such an opposition here, and use *dialogue system* to cover all these kinds of systems; the defining property here is that it is an (artificial) system that can enter into and hold some, perhaps limited, but in any case sustained form of (in the prototypical case) language-based interaction in real-time with a human. I will argue that for our purposes there are more commonalities between these different kinds of systems than is usually assumed, and that even the humblest kind of system (voice-only, not embodied) has to answer challenges that, depending on how and with which focus they are answered, can turn it into a cognitive model of an interesting type.

Now, what kind of analysis can dialogue systems offer, and of what? Let's first look at the task environment in which a dialogue agent finds itself. The information-processing task it needs to address is the quite substantial one of understanding language, and possibly a part of the world the conversation is about, well-enough to come up with a reaction, possibly in language as well, that is appropriate. (Note the restriction on *well-enough*; I will come back to this later.) This is the first step where dialogue systems can be usefully employed in cognitive modelling: building such a system forces one to precisely specify the task environment for (a particular setting of a) dialogue and the phenomenon of interest.

Given a particular conversational competence of interest (e.g., fast reaction times in turn-taking; more on possible modeling targets below in Section 3.2), a dialogue system can make, by embodying a computational model of them, theories of this competence testable. This property of making the predictions of a theory testable is something that dialogue systems of course share with any kind of computational model (for that is what dialogue systems are, to finally relate the discussion here to the previous section) in the cognitive sciences. However, they do this in an unusual way, by exposing themselves on-line to the situation type they are meant to model. With respect to the task of language processing, dialogue systems are *whole-agent models*: they need to say something about all levels of language processing (however many one assumes), from perceiving through understanding to generating it. This contrasts with the way for example theories of reading time are evaluated, namely against pre-collected

Computational Theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and the output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

Figure 1: The three levels of analysis of information processing tasks of (Marr, 1982)

corpus data (see e.g. (Lewis and Vasishth, 2005)); these are what could be called *sub-module models*.

For us, this property of being a *whole-agent model* is the 'unique selling proposition' of dialogue systems as *tools-for-understanding*. As complete models (w.r.t. a certain ability, and other constraints that will be discussed presently) of the agent-type they are meant to model, they have to produce a much wider range of behaviours than sub-module models, and have to be explicit about how these behaviours arise from that of the sub-modules (assuming that they do have discernible sub-modules). This is a challenge that can hardly be addressed otherwise, as (Marr, 1982) noted: "Almost never can a complex system of any kind be understood as a simple extrapolation from the properties of its elementary components".

It's not only the range of modelled behaviour where dialogue systems can have an advantage over off-line models, though. The kind of phenomena that seem to be promising goals for tackling in a dialogue system understood as cognitive model (see next section) also seem hard to model and evaluate otherwise. Decisions of an agent in a dialogue (the a_t from Section 2.1) typically have delayed rewards (how good was the conversation), and complete models of the world (that is, models of how the actions of the agent change the state of the world, $P(s_t|s_{t-1}, a_t)$, and of how the world is perceived, $P(o_t|s_t)$) are generally not available and, given the size of the state space, hard to learn from data—all of which suggests interactive evaluation as a strategy that is more promising than for example trying to reproduce a gold-standard from a corpus.[3]

The on-line nature of this interaction finally makes dialogue systems an ideal tool for exploring ideas from another recent approach within the cognitive sciences: *situated* or *embodied cognition*: "the theory of situated cognition [...] claims that every human thought and action is adapted to the environment, that is *situated*, because what people *perceive*, how they *conceive of their activity*, and what they *physically do* develop together." (Clancey, 1997, p.1). On-line interactions with dialogue systems inevitably happen in contexts, in situations, embedded at the very least in time, if not in space, and the systems need to address such situational features.

Let's wrap up the discussion of which of Marr's levels dialogue systems cover. As elaborated above, dialogue systems clearly represent a computational analysis: they contain a specification of what it is that is being computed, what the components of that computation are, and what the goals are. They are also by definition *implemented*—although most dialogue systems do not make any claims about the cognitive plausibility of the representations and algorithms they use. Lastly, most likely dialogue systems will not any time soon be able to tell us anything about the physical realisation of conversational skills, and hence aren't models on the physical level.

This then concludes this section: in the view proposed here, Dialogue Systems are situated, implemented *whole-agent models* of human language processing capabilities, and are as such computational cognitive models, perhaps with partial claims to representational and algorithmic realism as well.

3 Methodology and Domains

3.1 Of Robotic Bees and Conversational Agents

In (Michelsen et al., 1992), an experiment is described that represents the culmination of years of research on communication among honeybees: To test their understanding of the communication methods used by honeybees, the researchers built a mechanical model of a forager bee, put

[3] Interestingly, in the line of research that uses Bayesian methods like Reinforcement Learning to solve Partially-Observable Markov Decision Processes (see Lemon and Pietquin (2007) for a recent overview), a middle position is taken: the systems learn by interacting with user models which generate the observations, and which in turn are learnt from data. In effect, this is what could be called a "semi-interactive" setting, where two implemented models converse which each other.

it in a typical communication situation (inside a beehive), and let it perform various forms of dances, implementing variants of the models of bee-communication which the researchers had previously built from observation. The effectiveness of the dances (and hence the adequacy of the theories) was then evaluated by the number of bees that as a result flew to the predicted (communicated) locations.

We envision a quite similar place for dialogue systems in the study of human communication, and a similar methodology: artificial agents embody a theory of communication, whose adequacy is evaluated through the reactions it provokes in a naturalistic setting. However, compared to the honeybee, human communicative situations are somewhat more varied, and there are interesting interactions between technical limitations on what can be computationally modelled and choice of situation. The appropriate methodology then looks more like the following: a) start from theory that says something about phenomenon you want to study; b) devise communicative setting that keeps this phenomenon as unrestricted as possible while restricting other aspects as much as possible; c) record humans in this setting; d) derive from this a more fine-grained model, which is e) implemented in computational model; f) evaluate the model not only for how well it reproduces the phenomenon but also for the reactions it provokes.[4] (In practice, of course several iterations of c) to f) may be necessary.)

We go through the most important steps in the following.

3.2 Choice of Setting

The processing of human language poses quite formidable technical challenges, and the extant realisations even only of the sub-modules typically seen to be involved in it (e.g., parsing, "understanding", generation) are miles away from achieving human-like performance. This seems to pose a problem: if the components are that bad, how can we expect the result of their connection to be anywhere near a usable model of human behaviour (as in, one that helps answer interesting questions)?

The answer is, we shouldn't. Or at least we shouldn't be expecting to be able to model *unrestricted, intelligent conversation*. It is unrealistic to expect dialogue systems to be able to model "intelligent conversation" *per se*, that is, to expect them to be able to give "intelligent" replies to all kinds of utterances. Luckily, there are two (not mutually exclusive) ways around this problem. One is to restrict the setting in such a way as to require "intelligent" (or, better, appropriate) replies only in a narrow domain that *can* be modelled. The other is to shift the focus to other features of dialogue: Dialogue is not just about saying and meaning the right things. It's also about saying the right things at the right moment, and about giving the right kinds of other, not directly task-related signals.

It seems then that, at least in the short term, the most promising areas for modeling in dialogue systems are not those of the dynamics of meaning in dialogue, but that of the dynamics of interaction (where it is an interesting open question as to how much these can be disassociated). To give a laundry list of possible areas in control of interactivity that come to mind: turn-taking, timely feedback, emotional feedback, alignment between conversants. Also promising seems the study of emergent behaviours, created by interactions (planned / controlled or not) of parallel processes.

When the phenomenon of interest is selected and explanatory theories are consulted or constructed, the next step is to devise a setting in which the model can be evaluated. The challenge I see here is to choose a situation that reduces as much as possible the demands on the technical components, while still being as much as possible ecologically valid. The goal here is to externalise and expose the limits that the system has (insofar as they aren't part of what one wants to study) and to turn them into constraints posed by the situation (task, setting). E.g., a dialogue system will have understanding problems (ASR, NLU), so it's a good idea to restrict the situation in such a way that the space of expected interactions gets smaller, and the restrictions are intuitively clear to the human interactant.

To give an example for such a strategy (although the authors do not explicitly phrase it like this): in (Skantze and Schlangen, 2009), a system is presented that investigates how human-like levels of interactivity / turn-taking speeds can be reached. To investigate this, the authors restrict the situation into which the system is put to dictation of number

[4] Steps c) to f) follow the methodology proposed by (Cassell, 2007) for the construction of Embodied Conversational Agents.

sequences. This is a task that is intuitively understandable to human conversation partners, while making technical tasks that are not the direct goal of the investigation easier. (ASR can be expected to perform better on such a limited vocabulary.)

A lot of the ingenuity of using dialogue systems to answer questions about human language use will lie in the choice of restricted, but understandable settings.

3.3 Operationalisation, Model Construction

Once the setting has been determined and the general predictions of the theory have been mapped to it, the next step is to operationalise the theory so that it can be modelled computationally. Using the vocabulary introduced above, the task is to determine the range of actions that the system is meant to be able to take, the observations that are to be expected, and the state of the world that is to be tracked. (An additional detail is whether uncertainty about any of these elements should be modelled as well.)

Forcing explicitness at this step already is something that dialogue systems can contribute to the study of human language use. A functioning computational model of an ability (say, turn-taking) shows at least that the information given to it (say, word sequences and prosodic information) contains enough information to solve the computational problem.

In most current dialogue systems, the function from observations to actions is specified procedurally, as the outcome of the combination (in a pipeline, or partially parallel) of various processing modules. This reflects on the one hand what is seen as the structure of the problem—linguistics has traditionally separated the task of language understanding into the "modules" of syntax / parsing, semantics / interpretation and pragmatics / understanding—and on the other hand simply good software engineering practice. It also allows a more tentative approach, where less needs to be explicitly stated about the structure of the problem than what would be needed in a purely rational approach. (This of course can also be viewed as a downside of this approach.) Finally, as briefly mentioned above, it often is hard to get data from which free parameters of a rational model could be learned, and so analytical models with symbolic rules provide more control over the algorithm.[5]

It should be noted here that for the level of computational modeling, none of these differences matter. What matters here is a clear understanding of the problem; rational or probabilistic models perhaps have an advantage here because they enforce a clearer statement. If one puts weight on differences in processing mode, one starts to enter the algorithmic / representational level; for this to matter with respect to the modeling task, one would then need to claim realism for one way of processing or the other. Here again dialogue systems promise to be a useful tool, by making testable claims of advantages of different implementation methods.

The goal of studying human communication by means of computational modeling also gives the system designer the freedom to not fully implement those processing modules that aren't meant to be part of the model. For example, if the aim of the model is the study of discourse structure, and logical forms are required as input of the sub-module which is being tested, one could try a setting where a human "wizard" (Wooffitt et al., 1997) is in the loop—as long as this doesn't change the interactional dynamics one is interested in. Alternatively, an "oracle" could be employed: in a setting where what the human user will talk about is known in advance, for example because the user is asked to perform certain tasks, this information can be given to some modules of the system (unbeknown to the user) like reference resolver, speech recognition etc. Or, a system that is meant to model interaction features can use ELIZA-like techniques for content-management. (Cf. the discussion of "micro-domains" in (Edlund et al., 2008); more on this below.)

3.4 Evaluation

The final step is to evaluate the system for how well it does its job of modeling the phenomenon (and, more generally, of being 'human-like'). Evaluating dialogue systems is a difficult business, as has often been discussed (Walker et al., 1998; Edlund et al., 2008). The behaviour of a dialogue system is the result of the combination of many modules, and it is often difficult to ascertain which module's performance contributed what—asking the human users directly will often not give meaningful results.

[5] But see (Miller et al., 1996; Lemon and Pietquin, 2007; Schuler et al., 2009 in press) for some attempts at (partially) non-modular, probabilistic systems.

For using dialogue systems as *tools-for-understanding*, we see three basic ways for evaluation (which can be used together): First, if one has an objective measure of the modelled phenomenon available, one can treat the resulting interactions of human subjects and dialogue system as a corpus, and can compare the relevant measures in this corpus with measures of corpora of human–human interaction. Second, one can use subjective measures (user questionnaires) to evaluate the impression the system made. If one want to avoid asking directly for the feature one wants to evaluate, an indirect approach can be chosen where the evaluation question is held constant ("did you find the interaction similar to one with a human partner?"), but the system is varied along the interesting dimension (i.e., is intentionally 'disabled' wrt. the modelled phenomenon). Third, one can play or show the finished interactions to other experiment participants and let them evaluate the naturalness (a so-called "overhearer evaluation", (Whittaker and Walker, 2006)).[6]

4 Possible Objections

"Creating a human-like dialogue system means creating an Artificial Intelligence, and creating an Artificial Intelligence is impossible!"[7]

There are two parts to this objection. We'll deal with the last one first. Is creating an AI possible?

The criticism in (Larsson, 2005), if I understand it correctly, seems to turn on the assumption (following Dreyfus (1992)) that "the background [necessary for understanding human language] is not formalizable". The claim is that this applies both to attempts at explicitly formalising such background (e.g., using databases of facts and logical calculi to reason over them) as well as to learning approaches, and that from this observation it follows that "computers will never achieve human-level language understanding". While the position I've been advocating here does not require any claim about the possibility of human-level language understanding (more on this in a minute), I'd still like to note that I do not find the conclusion compelling.

The basis of the criticism seems to be the symbol grounding problem (see e.g. (Harnard, 1990)), i.e. the problem of providing abstract symbols with external, real-world meaning. In a quite sweeping manner, (Larsson, 2005) sets the bar for entry into the club of grounded beings high, and counts among the experiences that are required for understanding human language "being born by parents, going through childhood and adolescence and growing up and learning personal responsibility, social interaction". I do not see how a convincing in-principle argument can be formed along these lines. Ultimately, this seems to me an empirical question, and, *pace* (Wittgenstein, 1984 1953), I'd wait until I encounter a talking lion before I conclude whether I understand it or not.

Which brings me to the first part of the objection. Does the question whether building a (human-level) AI is possible even matter? Clearly, free conversation requires intelligence. Turing (1950) famously proposed a conversational deception test (am I talking to man or machine?) as a test for intelligence. But, as discussed above, human language use is not restricted to holding free conversation (and convincing the conversational partner one is human)—language is also used in other settings, and there are other competences that can be dissociated from this, and can be studied and modelled independently.[8,9]

Evaluation of these competences then amounts to running what could be called *Particularised Turing Tests*: Can the system convince the user that it is (like) a human operating under some, possibly relatively strict, constraints? An example could be a setting where the conversational partner is only allowed to ask questions. Do the utterances still come with a good timing? (The evaluation of course does not have to be Turing test-style, ie. as deception; see above for evaluation methodology.) (Edlund et al., 2008) call such settings "micro-domains", and specify as evaluation goal whether the system can be taken "for a human by *some person*, under *some set of circumstances*".

"Cognitive Science is about making predictions, not engineering systems. Building dialogue systems is an engineering task."

While the spoken dialogue systems technology is far away from providing standard environments

[6](Cassell, 2007) provides interesting anecdotal evidence of the use of this technique.

[7]A version of this objection has recently been raised in this forum (Larsson, 2005), and so we discuss is a bit more extensively here.

[8]To Larsson's (2005) credit, this is acknowledged in his criticism.

[9]Cf. the Practical Dialogue Hypothesis in (Allen et al., 2000): "The conversational competence required for practical dialogues, while still complex, is significantly simpler to achieve than general conversational competence."

like SOAR (Laird et al., 1987), components for example for ASR (e.g., Sphinx4, (Walker et al., 2004)) and dialogue managers (TRINDIkit, (Larsson and Traum, 2000)) are freely available. It is however true that considerable effort has to be spent on forming out of such components running systems into which one can build the models that are the primary interest. This can only get better if research groups start to share resources on a larger scale. Efforts to achieve this are currently underway (e.g., resources registry organised by SIGdial).

"You end up with bad cognitive science (too many compromises to get it to work at all) and bad engineering (too simple / useless domain)"

This is a serious objection. Attempting to use dialogue systems technology, which is still quite immature, can lead to making many compromises to just getting some form of reliable behaviour at all out the system. There is a danger of landing in a no-mans land, building a system that is neither particularly helpful in understanding the problems faced by human language processors or advances the state of technology. It is my opinion however that this can be avoided, and the methodology sketched above can help towards doing so.

"You need at the very least eyes, arms and legs to be cognitively plausible."

This is a (slightly caricaturising) summary of the central tenet of embodied cognition (Anderson, 2003). As mentioned above, I see dialogue systems as in any case *situated*, as they function in the same temporal environment as their conversation partner. When it comes to dealing with content, I am sympathetic with the view that grounding of symbols in percepts is a useful approach; however, as detailed above, not all of cognition having to do with language use is about content.

"People interact differently with machines and with humans, so machines have different computational problem to solve."

While there is evidence for the first part of the objection (Fischer, 2006), this also seem to depend on the metaphor with which human users enter into the interaction (Edlund et al., 2008). Moreover, in any case it is unlikely that human language users are even flexible enough to produce a *fundamentally* different kind of behaviour towards artificial conversational agents. The objection does however point out that it is important to frame the situation in which the model is evaluated carefully.

5 Dialogue Systems as Cognitive Models and as Computer Interfaces

Both Pieraccini and Huerta (2005) and Larsson (2005) point out that what we've called the *tool-for-understanding* and the *getting things done* approaches are complementary. In what sense, though? First, the differences. The directions answer to different constraints, to differences in what the free variables are. For cognitive models, the goal has to be human-like performance (wrt. the phenomenon being modelled), for practical system, the *primary* goal has to be efficiency and effectiveness wrt. to the task—human-likeness may or may not be a useful secondary goal. Consequently, the modeler in the *tool-for-understanding* view is free to choose a domain that lends itself best to an as-isolated-as-possible study of a phenomenon (see Section 3), while a researcher or practitioner building an applied system is free to implement behaviours that do not appear at all human-like.

So much for the differences. A common interest of course is to build components that help with language processing. Good speech recognition for example is as much a precondition for convincing computational models of language use as it is one for good practical systems. The overlap goes further, though. The already briefly mentioned work on POMDPs (Lemon and Pietquin, 2007) for example is, although being pursued more from an applied perspective, highly interesting also from a cognitive modeling perspective, as it uses techniques that can guarantee optimal computations.[10]

To conclude this section, I'd like to propose, with (Larsson, 2005), that "it would be good practice to explicitly state what the goals of a certain piece of research are", namely whether one wants to investigate human language use, using dialogue systems as a tool, or whether one wants to improve human–computer interaction.

6 Conclusions

In this paper, I have discussed the potential and possible problems of using spoken dialogue systems (ecumenically understood as all kinds of ar-

[10]Interestingly, there is some reservation against such methods from a commercial perspective (Paek and Pieraccini, 2008), where the additional constraint of provability of dialogue strategies seems to be important for customers who employ such systems.

tificial systems that can interact via spoken natural language) as models of (certain aspects of) human cognition. I have sketched a methodology for doing so, proposing that the main use of dialogue systems for now lies in how they can help being more explicit about one structures the tasks.

The models that can be built at the moment are rather crude and limited, and necessarily containing many simplifications. The hope is that combined efforts on practical systems and on systems built as *tools-for-understanding* can improve both kinds of systems, and help advance our understanding of human language use.

Acknowledgments: The author gratefully acknowledges support by DFG (Emmy Noether Programme).

References

James F. Allen, Lenhart K. Shubert, George Ferguson, Peter Heeman, Chung Hee Hwang, Tsuneaki Kato, Marc Light, Nathaniel G. Martin, Bradford W. Miller, Massimo Poesio, and David R. Traum. 1995. The TRAINS project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI*, 7:7–48.

James F. Allen, Donna Byron, M. Dzikovska, George Ferguson, L. Galescu, and A. Stent. 2000. An architecture for a generic dialogue shell. *Natural Language Engineering*, 6(3).

John R. Anderson. 1991. The place of cognitive architecture in a rational analysis. In K. van Lehn, editor, *Architectures for Intelligence*, chapter 1, pages 1–24. Lawrence Erlbaum Associates, Hillsdale, N.J., USA.

Michael L. Anderson. 2003. Embodied cognition: A field guide. *Artificial Intelligence*, 149:91–130.

Justine Cassell. 2007. Body language: Lessons from the near-human. In Jessica Riskin, editor, *Genesis Redux: Essays in the History and Philosophy of Artificial Life*. Chicago University Press, Chicago, USA.

Nick Chater and Mike Oaksford, editors. 2008. *The Probabilistic Mind: Prospects for a Bayesian cognitive science*. Oxford University Press, Oxford, UK.

William J. Clancey. 1997. *Situated Cognition: On Human Knowledge and Computer Representation*. Cambridge University Press, Cambridge, UK.

David DeVault and Matthew Stone. 2009. Learning to interpret utterances using dialogue history. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 184–192, Athens, Greece, March. Association for Computational Linguistics.

Hubert Dreyfus. 1992. *What computers still can't do*. MIT Press, Boston, Massachusetts, USA.

Jens Edlund, Joakim Gustafson, Mattias Heldner, and Anna Hjalmarsson. 2008. Towards human-like spoken dialogue systems. *Speech Communication*, 50:630–645.

Kerstin Fischer. 2006. *What Computer Talk Is and Is not: Human-Computer Conversation as Intercultural Communication*. Linguistics – Computational Linguistics. AQ-Verlag, Saarbrücken, Germany.

Stevan Harnard. 1990. The symbol grounding problem. *Physica D*, 42:335–346.

Michael I. Jordan and Stuart Russell. 1999. Computational intelligence. In *The MIT Encyclopedia of the Cognitive Sciences*, pages lxxvi–xc. MIT Press, Cambridge, Massachusetts, USA.

John Laird, Allan Newell, and P. Rosenbloom. 1987. SOAR: An architecture for general intelligence. *Artificial Intelligence*, 33(1):1–64.

Staffan Larsson and David R. Traum. 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6(3–4).

Staffan Larsson. 2005. Dialogue systems: Simulations or interfaces. In *Proceedings of* DIALOR*, the 9th Workshop on the Semantics and Pragmatics of Dialogue*, Nancy, France.

Oliver Lemon and Olivier Pietquin. 2007. Machine learning for spoken dialogue systems. In *Proceedings of Interspeech 2007*, Antwerp, Belgium.

R.L. Lewis and S. Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3):375–419.

David Marr. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, San Francisco, USA.

Axel Michelsen, B. B. Anderson, J. Storm, W. H. Kirchner, and M. Lindauer. 1992. How honeybees perceive communication dances, studied by means of a mechanical model. *Behav. Ecol. Sociobiol.*, 30:143–150.

Scott Miller, David Stallard, Robert Brobow, and Richard Schwartz. 1996. A fully statistical approach to natural language interfaces. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, California, USA.

Tim Paek and Roberto Pieraccini. 2008. Automating spoken dialogue management design using machine learning: An industry perspective. *Speech Communication*, 50:716–729.

Roberto Pieraccini and Juan Huerta. 2005. Where do we go from here? research and commercial spoken dialog systems. In *Proceedings of the 6th SIGdial workshop on Discourse and Dialogue*, Lisbon, Portugal, September.

William Schuler, Stephen Wu, and Lane Schwartz. 2009 in press. A framework for fast incremental interpretation during speech decoding. *Computational Linguistics*.

Gabriel Skantze and David Schlangen. 2009. Incremental dialogue processing in a micro-domain. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009)*, pages 745–753, Athens, Greece, March.

Alan Turing. 1950. Computing machinery and intelligence. *Mind*, 59:433–460.

Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella. 1998. Evaluating spoken dialogue agents with PARADISE: Two case studies. *Computer Speech and Language*, 12(3).

Willi Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, Evandro Gouvea, Peter Wolf, and Joe Woelfel. 2004. Sphinx-4: A flexible open source framework for speech recognition. Technical Report SMLI TR2004-0811, Sun Microsystems Inc.

Steve Whittaker and Marilyn Walker. 2006. Evaluating dialogue strategies in multimodal dialogue systems. In Wolfgang Minker, D. Bühler, and Laila Dybkjaer, editors, *Spoken Multimodal Human–Computer Dialogue in Mobile Environments*. Springer, Den Haag, The Netherlands.

Ludwig Wittgenstein. 1984 [1953]. *Tractatus Logicus Philosophicus und Philosophische Untersuchungen*, volume 1 of *Werkausgabe*. Suhrkamp, Frankfurt am Main.

R. Wooffitt, N.M. Fraser, N. Gilber, and S. McGlashan. 1997. *Humans, Computers and Wizards*. Routledge, London and New York.