Dialogue Systems: Simulations or Interfaces?

Staffan Larsson
Dept. of linguistics
Göteborg University
SE 40530 Göteborg, Sweden
sl@ling.gu.se

Abstract

This paper raises the question of the aim and scope of formal research on dialogue. Two possible answers are distinguished – the "engineering" and the "simulation" view – and an argument against the soundness of the "simulation" position is reviewed. This argument centres on the (im)possibility of formalising the context (or "background") needed for human-level language understanding. This argument is then applied to formal dialogue research and some consequences are discussed.

1 Introduction

Although perhaps nowadays many researchers would be wary of subscribing to the view that a complete simulation of human language use is possible, the precise extent to which this goal is feasible (and desirable) is still an open question. A premise of this paper is that this is an important issue to discuss, and that such a discussion could be useful as a backdrop for formulating goals and methods for research on the formal semantics and pragmatics of dialogue.

In this paper, I raise the question of the aim and scope of research on dialogue systems and the formal and computational semantics and pragmatics of dialogue. (I will refer to this area of research as "formal dialogue" research.) I distinguish two possible answers – the "engineering" (or "interface") view and the "simulation" view – representing the most extreme positions taken in response to this question. I then review an argument against the soundness of the "simulation" position, in order to give an impression of the deep difficulties involved in achieving this goal. This argument centres on the (im)possibility of formalising the context (or "background") needed for human-level language understanding.

The contribution of the present paper is the explicit application of this argument to formal dialogue research and an attempt to draw out some consequences of the argument for this area of research. I argue that an intermediate position closer to the "engineering" view on formal dialogue research is both more useful and realistic. However, knowledge of human language use (both formal and informal) is still essential in this endeavour. A further important consequence of the argument is that since the "simulation" and "interface" research programs are in fact very different, it is important to be clear about the goal in any given piece of work in formal dialogue research.

2 Engineering vs. Simulation

The first view (the "engineering" position) claims that the purpose of formal dialogue research is ultimately one of interface engineering; to enable the building of better human-computer interfaces by incorporating (spoken) dialogue. The second answer (the "simulation" position) claims that the ultimate goal is a complete computational (implementable) theory of human language use and understanding. In reality, there is of course a continuum where individual researchers may take intermediate positions, take different positions depending on the situation, and/or assume that both goals converge and so there is no reason to take any position (which is, of course, itself a position). For example, one intermediate position might be to regard formal semantics and pragmatics as capturing (although in a more or less simplified manner) some aspects of human language use while deliberately ignoring or oversimplifying other aspects, and to regard dialogue systems as a possible area of application for such theories. Still, the issue remains as to what the ultimate goal of the research is.

3 Dialogue systems as interfaces

There is, I believe, a consensus in the formal dialogue research community that (spoken) dialogue has the potential of vastly improving on, or even replacing, available human-computer interface technology. There are good reasons for this optimism, as spoken dialogue is perhaps the most natural way for humans to communicate. As technologies become more complex, previous interfaces such as the command-line or menu-based graphical interfaces become increasingly unwieldy and impractical, and an interface based on the metaphor of intelligent conversational agents becomes increasingly attractive.

A common idea in formal research on dialogue is that there is a extensive, if not complete, overlap between research on human language use and research on conversational interfaces. To build good conversational interfaces it is important to develop, extend, formalise and implement theories of human language use. A very influential way of thinking about this overlap is the idea that a dialogue system should, as far as possible, be a *simulation of human language use*.

4 Dialogue systems as simulations

Can a machine be intelligent? Turing offered an operational definition of intelligence in the form of a test, which goes roughly like this: Test person A has a dialogue (via a text terminal) with B. A:s goal is to decide whether B is a human or a machine. If B is a machine and manages to deceive A that B is a human, B should be regarded as intelligent and able to think.

According to the Turing test, human intelligent behaviour is equivalent to the ability to carry out a dialogue using natural language. This means that in order to make a computer use natural language in the same way and on the same level as a human, it needs to be endowed with human-level intelligence. Interestingly, this also means that research on semantics and pragmatics of natural language has a central role in AI as a whole. In fact, if one takes a simulation view on formal dialogue research, this field becomes in a way equivalent to AI. Below, I will review an argument against AI, and attempt apply it more explicitly to research on dialogue systems and formal semantics and pragmatics.

4.1 GOFAI and formal research on dialogue

A lot of formal research on dialogue has, by way of inheritance or common ancestry, some central ideas in common with the classical AI approaches (sometimes referred to as Good Old-Fashioned AI, or GOFAI (Haugeland, 1985)). For example, the Information

State Update approach to dialogue management (Traum and Larsson, 2003) has a lot in common with GOFAI approaches such as SOAR (Laird et al., 1987). Symbolic representation and symbol manipulation remain important cornerstones in the way that problems are formulated and in the form of the solutions given. Starting out from the assumption that sentences in natural language can be given a formal semantics, the realisation that context plays a central part in language use has led to the idea of formalising the context so that it can be related to formal semantic representations of sentences and utterances. One reason for the use of formal techniques is simply that, so far, representation and symbol manipulation it seems to be the most (or even the only) workable method for dealing with many of the complex problems of natural language dialogue, e.g. ellipsis resolution, pronoun resolution, dialogue act recognition, keeping track of multiple topics, etc..

Often, computational dialogue researchers implement their theories either in limited toy examples, or as semi-functional dialogue system interfaces for small domains [1]. Dialogue systems based on symbolic computation definitely appears to be useful for improving on current human-computer interfaces, although only a major breakthrough of natural language dialogue interfaces would prove this conclusively. But is it also a step on the way towards human-level natural language understanding in computers?

[1] A recent paper (Bos, 2005) boldly claims to show that "it is possible to have a robust and wide-coverage system that generates semantic interpretations with relevant background knowledge from texts and perform first-order inferences on the result." However, the correctness (accuracy and adequacy) of these resulting representations have not yet been evaluated. As I understand Dreyfus' argument it would predict that the correctness would be low in most domains, although some success is possible in systematic domains (see Section 5.1).

4.2 Arguments against GOFAI

The position of so-called "weak AI" is, roughly, that computers can be made to act as if they were intelligent (Russell S, 1995). Independently of any "strong AI" claims as to whether such a computer would also be conscious, Dreyfus (1992) and others (e.g. Winograd and Flores (1987)) have put forward arguments against the possibility of weak AI, based on the philosophies of Heidegger, Merleau-Ponty, and the later Wittgenstein. As these arguments centre on the possibility of human-level understanding of language in computers, they are also very relevant to the present discussion. This section briefly reviews Dreyfus' arguments; unfortunately, space restrictions make it hard to do justice to the argumentation. For the full account, see Dreyfus (1992).

Some well-known problems in GOFAI are computational complexity of logical inference in real-time resource-bounded applications, planning for conjunctive goals, plan recognition, incompleteness of general FOL reasoning (not to mention modal logic), and the frame problem (Haugeland, 1987). However, as humans we don't tend to encounter these problems in our everyday life (unless, of course, we happen to be AI researchers). Dreyfus asks rhetorically whether it is is possible that all these problems have a common cause? Well, they all seem to be related to symbolic representations and symbol manipulation. The idea that understanding and thinking is forming and using symbolic representations is an old one, going back at least to Descartes[2] and reformulated in (Newell and Simon, 1963) as the "physical symbol hypothesis". According to this idea, intelligent behaviour can be captured by a system that reasons logically from a set of formal and context-independent facts and rules[3].

[2] Dreyfus argues that the idea of formalising human reasoning goes back at least to Plato.

[3] Although facts in themselves may (purport to) represent

Against this, Dreyfus argues that human behaviour is essentially non-formal. Human behaviour based on our everyday common-sense background understanding, which allows us to experience what is currently relevant, deal with things and people in everyday situations, and understand natural language. The background involves (among other things) utterance situation, ongoing activities, relevant institutions, and cultural settings. In its widest sense, the background involves all of human culture and experience as it is passed down through generations in social interaction. Dreyfus argues that the background has the form of dispositions, or informal know-how. It is thus a form of skill rather than propositional knowing-that – inarticulate, and to some extent pre-conceptual.

To achieve GOFAI, this know-how, along with the interests, feelings, motivations, and bodily capacities that go to make a human being, would have to be conveyed to the computer as knowledge in the form of a huge and complex belief system. Indeed, work in this direction has been going on for several years, e.g. in the CYC project (Lenat and Guha, 1990). Dreyfus argues, however, that *the background cannot be formalised*; there are no reasons to think that humans represent and manipulate the background explicitly or that this is at all possible even in principle. To quote from Dreyfus (1992), p. 3: "...understanding requires giving the computer a background of common sense that adult human beings have by virtue of having bodies, interacting skillfully with the material world, and being trained into a culture.". This background enables humans to, among other things, skillfully cope with changing events and motivations, project understanding onto new situations, and understand social innovations – someone may do something that has not so far counted as appropriate, and have it recognized in retrospect as having been just the right thing to do

Even so, there is a grain of truth to the information-processing idea. When something goes wrong – when there is a *breakdown* in some activity – we need to reflect and reason, and may have to learn and apply formal rules. However, it is a mistake to read these rules back into the normal situation and appeal to such rules for a causal explanation of skillful behaviour. Similarly, when learning new skills we might start from a set of rules and facts, but as we progress from novice to expert, the rules are replaced by embodied skills.

4.3 Non-symbolic approaches to AI and dialogue

According to Dreyfus, since around 1986 GOFAI has become less popular, partly in response to arguments from critics such as Dreyfus. A widely-used textbook on AI acknowledges admits that "[m]any of the issues Dreyfus discusses (...) are now widely accepted as important aspects of intelligent agent design." (Russell S, 1995). Instead, there has been an increasing focus on non-symbolic or semi-symbolic methods such as connectionism, embodied interactive automata, reinforcement learning, probabilistic methods, etc.. Mirroring this move in cognitive science is an increased focus in computational linguistics, including formal dialogue research, on semi-symbolic statistical and machine-learning methods.

Space restrictions prohibit a thorough discussion of whether non-symbolic methods can be used to overcome the problem of equipping computers with the background necessary for human-level language understanding. Suffice to say that Dreyfus argues (convincingly, in my view) that non-symbolic approaches to AI face the same basic problem as the symbolic approach. True, non-

context, and rules may make reference to such facts, the facts and rules themselves are represented in a formal language or programming language which does not depend on context for its interpretation.

symbolic systems do not themselves contain a formal description of background. However, they cannot be built and trained without a pre-existing formalisation of background knowledge.

To put it very briefly, the reason is that even these approaches require a formalised context in order to set up the training data in a way that will allow a system to learn anything useful from it. This requires that humans interpret the context in terms of its relevant features before it can be fed to the computer. To quote from a recent conference call[4]:

> As experience with machine learning for solving natural language processing tasks accumulates in the field, practitioners are finding that feature engineering is as critical as the choice of machine learning algorithm, if not more so. Feature design, feature selection, and feature impact ... significantly affect the performance of systems and deserve greater attention.

This process of "feature engineering" is far from an innocent "preparation" of data; rather, it is a crucial step of *pre-digesting* the data by noting the relevant aspects of a situation to a problem at hand and embodying this interpretation in a formal description that the computer can then manipulate. The quote above indicates that there seems to be a growing realisation within the AI community that "feature engineering" is crucial for natural language processing in computers.

The ability to see the relevant features of a situation is not present in computers, Dreyfus argues, since it crucially requires a commonsense background. So, one might wonder, how do humans manage to learn this background? As already indicated in the quote above, they are able to do so by virtue of having bodies, interacting skillfully with the material world, and being trained into a culture. Language is, simply, very deeply interconnected with human life. Unless we are able to build computers which have (human or human-like) bodies, and which are trained into a culture through social practices of human society (involving being born by parents, going through childhood and adolescence and growing up and learning personal responsibility, social interaction, making friends, and establishing an identity, and all the other things that make up human life), the argument implies, no machine will ever pass the Turing test[5].

It must be stressed that this is not a "knock-down argument" proving conclusively that weak AI is impossible; no such claims are made by Dreyfus. For me personally, it served to point out that achieving human-level language understanding in computers might be much harder than I had previously thought, and that the research methods involved in pursuing this goal may be quite different from the methods appropriate for the design of dialogue systems as human-computer interfaces.

5 Formal dialogue research and dialogue systems design

If we accept the argument that "the background is not formalizable" and that computers will never achieve human-level language understanding, does it follow that formal and computational research on dialogue and dialogue systems is useless? Of course not; it provides (as already mentioned) a great potential for improving on human-computer interaction. But granted this, has theories of human language use now been shown to be of no use to research on human-computer dialogue? Again, of course not. For one thing, if we want dialogue systems that are reasonably human-like in their behaviour, these systems will need to be designed on the basis of theories of human language. But this does not require that these theories have to be formal descriptions on human language use and

[4] `http://research.microsoft.com/~ringger/FeatureEngineeringWorkshop/`

[5] Indeed, it could be argued that the power of the Turing test rests on this intuition that the ability to carry on a dialogue in natural language truly requires *human* intelligence.

cognition, nor of implementations of them as (even partial) simulations. Instead, we may use these theories as providing important clues about how to best build dialogue systems. Firstly, we may observe regularities in dialogue that can serve as the basis for formal representations. Second, non-formal theories of those aspects of language use which resist formalisation can be used as a basis for design of aspects of dialogue systems that do not need to be modelled by the system itself.

5.1 Formal theories as systematic domains

Arguing against the possibility of human-level intelligence and language understanding by computers (along similar lines as Dreyfus), Winograd points out that computers are nevertheless useful tools in areas of human activity where formal representation and manipulation is crucial, e.g. word processing. In addition, many practical AI-style applications do not require human-level understanding of language. In such cases, it is possible to develop useful systems that have a limited repertoire of linguistic interaction. There are regularities in conversational behaviour ("domains of recurrence"), and that on the basis of such regularities it is possible for e.g. a researcher to create[6] so-called *systematic domains*. That is, a set formal representations that can be used in a system and that embodies the researcher's interpretation of the situation in which the system will function.

Note that providing formal rules for describing behaviour does not necessarily imply that similar rules are represented in humans. If we accept that human behaviour is essentially non-formalizable, formal rules will always be, at best, rough representational approximations of the non-representational know-how embodied in humans[7].

[6]Winograd stresses that this is a creative process, rather than one of mere observation.

[7]Compare the case where a statistical speech recognition grammar is trained on the output of a formal grammar. The SLM can then be subjected to machine learning which will subtly modify its behaviour in ways that could not be expressed in the rules of the original grammar. The difference is only that in attempting to formally describe language use, we are abstracting the hard-edged rules from embodied behaviour, rather than starting with the rules. Humans may have learned their behaviour with or without starting from explicit rules; however, human behaviour is always shaped by biological factors and social interactions that are not available to computers for reasons already discussed.

Semantics[8] is not a focus of Winograd's formal analysis, presumably because Winograd believes that language understanding is not amenable to formal analysis (see also citewindograd:fulcrum). However, even if one accepts the arguments such as those above, I believe that the idea of systematic domains also applies to semantics. That is, for certain "semantically systematic" task domains it is indeed possible to provide a formal semantics, e.g. in the form of a formal ontology and formal representations of utterance contents. Arguably, semantics is more closely related to specific activities than is pragmatics, since semantics involves the entities and relations which are relevant in a given activity or task domain. This means that the question of whether a task domain can be usefully captured in formal semantic must be answered for each task domain or task domain type individually.

5.2 Non-formal theory and dialogue systems design

As mentioned above, non-formal theories of those aspects of language use which resist formalisation can be used as a basis for design of aspects of dialogue systems that do not need to be modelled by the system itself. For example, it is likely that any speech synthesizer voice has certain emotional or other cognitive connotations; it might sound silly, angry, etc.. It is extremely difficult, if not impossible, to design a completely neutral voice. However, if we have some idea of how different voices are perceived (or perhaps even how different

[8]I am not claiming that there is a strict division between semantics and pragmatics.

aspects of speech synthesis correlate with the perceived "personality" of the voice), we can use this (informal) knowledge to provide a dialogue system application with an appropriate voice for that application.

6 Conclusions

This paper has distinguished two extreme views on formal dialogue research; the "engineering" view and the "simulation view". On the basis of Dreyfus' criticism of AI, I have argued that the simulation view, at least in its most extreme form, is probably untenable as an explicit or implicit research goal. I have also argued, on the basis of Winograd's ideas, that formal dialogue research may nevertheless be useful for improving dialogue systems in limited domains and with limited linguistic capabilities. I have also suggested that formal theories of language use are limited in scope and should be complemented by non-formal theories in the design of dialogue system interfaces.

Domains of language use that may be susceptible to formalisation (i.e. creation of systematic domains) can be roughly divided into pragmatic and semantic domains. Pragmatic domains include e.g. aspects of dialogue management such as turntaking, feedback and grounding, referent resolution, and topic management and sequencing. Issues related to semantic domains concern e.g. application-specific ontologies and the fine-grainedness and expressivity of the formal semantic representation required for a domain or group of domains (e.g. database search, device programming, collaborative planning). The general issue of how to determine whether a task domain is amenable to formal semantic description is one that deserves to be further investigated, as well as the closely related issue of how to extract a formal description from available data of the domain, e.g. transcripts of dialogues. A third related issue is how to decide, for a given task domain, what level of sophistication is required by a formal semantic framework in order for it to be useful in that domain. In some domains, simple feature-value frames may be sufficient while others may require something along the lines of situation semantics, providing treatments of intensional contexts etc.[9]

The question is still open exactly how far it is possible to go in the formal description of phenomena related to language use, and the only way to find out is to by trial-and-error (i.e., research). I'm thus by no means arguing that one should stop trying to extend the coverage of formal semantics and pragmatics, rather that one might be well-advised to keep in mind the following points:

1. Formal theories of language use should be regarded as the result of a creative process of finding regularities in language use as a basis for the construction of formal representations that can be used in dialogue systems to open up new possibilities for human activity.

2. Even though some aspects of language use may indeed be susceptible to formal description, this does not mean that human language use actually relies on such formal descriptions represented in the brain or elsewhere.

3. Repeated failures to formally capture some aspect of human language may be due to the limits of formal theory when it comes to human language use, rather than to some aspect of the theory that just needs a little more tweaking.

Even though the arguments against AI cited above constitute, in my view, good reasons

[9] As a special case, when adapting a dialogue system to function as an interface to an existing program, there is already a formalised domain in the form of the actions, entities etc. of the existing interface. In such cases, it is usually sufficient with a very basic semantic formalism. In addition, existing interfaces such as menu-based GUIs can provide a readily available formalisation of useful conversational structures, e.g. by converting menu systems into dialogue plans.

to be very sceptical about the possibility of simulating human language use, it would certainly be premature to completely abandon this research[10]. However, as this (as has been argued above) constitutes a rather different project than that of building good interfaces and tools for systematic domains, it would be good practice to explicitly state what the goals of a certain piece of research are in case this is not obvious.

Being clear about this can also serve to motivate different research strategies and to estimate the validity of different types of argument used when presenting the research. For research taking the "engineering" view, methodologies should be concerned not primarily with how human cognition and language use works, but rather with designing and engineering of useful and flexible human-computer dialogue interfaces. If interface engineering is liberated from concerns related to simulation, it can instead be focused on the creation of new forms of human-computer (and computer-mediated) communication, adapted to and exploring the respective limitations and strengths of humans and computers. Of course, knowledge about human language use is relevant here as well (as a source of inspiration, if nothing else) but is not regarded as an end in itself.

References

Johan Bos. 2005. Towards wide-coverage semantic interpretation. In *Proceedings of Sixth International Workshop on Computational Semantics (IWCS-6)*.

Rudi Cilibrasi and Paul Vitanyi. 2005. Automatic meaning discovery using google. Technical report, University of Amsterdam, National ICT of Australia.

Hubert Dreyfus. 1992. *What computers still can't do*. The MIT Press.

John Haugeland. 1985. *Artificial intelligence: the very idea*. Massachusetts Institute of Technology, Cambridge, MA, USA.

John Haugeland. 1987. An overview of the frame problem. In Z. W. Pylyshyn, editor, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Theoretical Issues in Cognitive Science, pages 77–93. Ablex, Norwood, New Jersey.

J. Laird, A. Newell, and P. Rosenbloom. 1987. Soar: An architecture for general intelligence. *Artificial Intelligence*, 33(1):1–64.

D. Lenat and R.V. Guha. 1990. *Building Large Knowledge-Based Systems - Representation and Inference in the Cyc Project.* Addison-Wesley, Reading, Massachusetts.

Alan Newell and Herbert Simon. 1963. GPS: A program that simulates human thought. In Edward Feigenbaum and Jerome Feldman, editors, *Computers and Thought*, pages 279–293. McGraw-Hill.

Norvig P Russell S. 1995. *Artificial Intelligence: A Modern Approach*. Prentice Hall Series in Artificial Intelligence. Englewood Cliffs, New Jersey.

Luc Steels and Paul Vogt. 1997. Grounding adaptive language games in robotic agents. In C. Husbands and I. Harvey, editors, *Proceedings of the Fourth European Conference on Artificial Life*. The MIT Press.

David Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In Ronnie Smith and Jan Kuppevelt, editors, *Current and New Directions in Discourse & Dialogue*. Kluwer Academic Publishers.

T Winograd and F. Flores. 1987. *Understanding Computers and Cognition*. Addison-Wesley Professional.

Terry Winograd. 1985. Moving the semantic fulcrum. *Linguistics and Philosophy*, 8(1):91–104.

[10] Some recent developments in non-symbolic or semi-symbolic AI that can be seen as addressing the problems discussed above (e.g., overcoming the need for formal descriptions of context, produced by humans) include simulated evolution of linguistic communication between robots (Steels and Vogt, 1997), and mining for semantic relations on the web (Cilibrasi and Vitanyi, 2005). The former approach does not, however, aim to evolve *human* language, and there are good reasons for this (Winograd, 1985). The latter approach is certainly interesting, but there are reasons to be sceptical as to how far it can get, as it does not address many of the issues raised by Dreyfus, e.g. the importance of embodiment and social interaction.