Capturing the Implicit – *an iterative approach to enculturing artificial agents*

Peter Wallis and Bruce Edmonds

Centre for Policy Modelling
Manchester Metropolitan University
Manchester, United Kingdom
pwallis@acm.org, bruce@edmonds.name

Abstract. Artificial agents of many kinds increasingly intrude into the human sphere. SatNavs, help systems, automatic telephone answering systems, and even robotic vacuum cleaners are positioned to do more than exist on the side-lines as potential tools. These devices, intentionally or not, often act in a way that intrudes into our social life. Virtual assistants pop up offering help when an error is encountered, the robot vacuum cleaner starts to clean while one is having tea with the vicar, and automated call handling systems refuse to let you do what you want until you have answered a list of questions. This paper addresses the problem of how to produce artificial agents that are less socially inept. A distinction is drawn between things which are operationally available to us as human conversationalists and the things that are available to a third party (e.g. a scientists or engineer) in terms of an explicit explanation or representation. The former implies a detailed skill at recognising and negotiating the subtle and context-dependent rules of human social interaction, but this skill is largely unconscious – we do not know how we do it, in the sense of the later kind of understanding. The paper proposes a process that bootstraps an incomplete formal functional understanding of human social interaction via an iterative approach using interaction with a native. Each cycle of this iteration entering and correcting a narrative summary of what is happening in recordings of interactions with the automatic agent. This interaction is managed and guided through an "annotators' work bench" that uses the current functional understanding to highlight when user input is not consistent with the current understanding, suggesting alternatives and accepting new suggestions via a structured dialogue. This relies on the fact that people are much better at noticing when dialogue is "wrong" and in making alternate suggestions than theorising about social language use. This, we argue, would allow the iterative process to build up understanding and hence CA scripts that fit better within the human social world. Some preliminary work in this direction is described.

1 Introduction

This paper is focused upon computers that inhabit roles with human origin. In particular, computers that have to converse with people as social actors, in the course of their interactions with them. This is not the only sort of interface of course and some will argue that computers as we know them have perfectly satisfactory interfaces, e.g. those based on the notion that the computers are a tool facilitated by a physical analogue (e.g.

a desktop). However a "social stance" – considering computers as social actors – may allow for a new range of applications to emerge as well as giving new insights into human behaviour, in particular the current limitations of our models of this.

However, when computers are compelled to work as social actors – for example when they use language as the primary modality – they tend to fail grossly rather than in detail. Indeed people get so frustrated by computers that they often swear at them [1]. When someone swears at a carefully crafted chat-bot, the human is unlikely to have been upset by punctuation or a quirky use of pronouns. The challenge is that existing qualitative techniques are good at the detail, but can fail to find a bigger picture.

In this paper the focus is on performative language and builds on the findings of applied linguistics where the mechanism of language can be seen as part of the same spectrum of communicative acts ranging from "body language" to semiotics. However much of these communicative acts are learned in context with reference to their effect rather than a putative explicit meaning.

This is contrary to approach that characterises human actors as rational actors. Applied to language this motivates the characterisation that natural languages are a "fallen" version of something more pure – a messy version of First Order Predicate Calculus – where elements of the language can be associated with their separate meaning. This meaning-text model [2] has been largely rejected since the late 1980's but has a latent existence in the idea that it is possible to create sets of Dialogue Acts (DAs) that capture in some way the primitive concepts from which any conversation can be constructed. For a comprehensive description of the theory and lack thereof in this area, there are several papers by Eduard Hovy [3].

It is also going in a different direction to those focused on statistical and machine learning techniques [4] that treat mental attitudes as a "hidden state" that can be derived from corpora of human behaviour. The advantage of this approach to engineering dialogue systems is that we do not need to understand how language is used, the machine will figure it out for itself (as far as it is able). The challenge is the amount of training data required to cover all the necessary cases and, unlike a search engine where measured performance as low as 10% is useful, many errors social actors make are noticed and need to be dealt with. The assumption here is that we want to know more about the process of being a social actor, and know enough about it to be able to make a computer to do it with sufficient competency.

Rather this paper is predicated on the notion that there is a wealth of vague, implicit, context-dependent and often unconscious knowledge that is necessary for a social actor to successfully inhabit a society [5, 6], and to show how such knowledge might be incorporated into artificial agents. Such social knowledge is not immediately accessible to an engineer as explicit knowledge and so the classic "waterfall" model of software engineering, in which one starts by developing a detailed specification and follows up with a development phase, is inappropriate. Instead, a process of entity enculturation – learning how a CA should behave in context over a period of time – is required. Design plays a part, but has to be leveraged by a substantial subsequent iterative process of trial and repair [7]. This is not a "one off" method of making socially fluent agents, but a method of repeatedly: (1) analysing records of their interaction in situ, then (2) affecting a repair on the behaviour for this context. Thus, over time, embedding the agent into the culture it perforce inhabits.

The core of this approach is the leveraging of a common narrative understanding of interactions between people. In this non-scientists are asked to “tell the story” of how a particular situation came about with a conversational agent as a starting point in preparation for an iterative approach to repairing that agent. In subsequent iterations they may be asked to comment upon an existing narrative, possibly entering alternative descriptions at certain points. This interaction will be guided and constrained by a a developer’s workbench that allows someone to both “script” future dialogue and analyse recordings of past (real) dialogue with the machine by narrating the action, and would capture the *mechanism* by which we social actors decide what to say and when.

2 Contributory Threads

Given the nature of the proposal, and its contrary direction it is useful to trace the projects and results that have lead us in this direction.

2.1 The KT experiments

The KT experiments were a project to understand the issues and the potential for embodied conversational agents (ECA) acting as virtual assistants [8]. As part of that project, we conducted Wizard-of-Oz experiments, where a human covertly pretends to be the conversational agent conducting the conversations, followed by interviewing the wizard (KT) about her actions using a technique from applied psychology called Applied Cognitive Task Analysis (ACTA) [9]. The aim was to populate a model of KT doing the task, and then use that model to drive a virtual assistant performing the same task. The model was “folk psychological” in that it her beliefs, desires and other mental attitudes were used as theory to explain and identify the “causes” of her behaviour. For these experiments the task was simply to have staff call our agent when they wanted to use one of the cars from the Division’s car pool. Ultimately the task was a slot-filling task: specifying which car, who was using it and the time.

The relevant results were twofold. *Firstly*, that politeness was more important than getting the facts right. For various reasons KT’s “slot-fill rate” - how often she managed to identify a piece of information in the caller’s utterances and enter it in the appropriate slot - was just over 80%. A “fact error rate” of close to 20% might sound high but the point is *nobody minded* and, although we didn’t measure it, we expect nobody *noticed*. Why didn’t they mind or notice? Because of course KT would make appropriate apologies and gave explanations when she had forgotten what they said their phone number was or where they were going. What is more, looking at the length of utterances, it was easy to see how KT’s utterances could convey the same information in a more compact form. Grice’s Maxims would suggest shorter is better (a principle popular with call centre industry) but KT did not want to use shorter utterances because, from the interview process, it just wouldn’t be polite. As a scientist one might have theories about the concept of face [10], but KT’s seems to have some system for doing social interaction that uses politeness as an atomic concept. She had not read Brown and Levinson's book [10] and didn't need to.

Secondly, it turned out that interviewing people about their everyday behaviour is problematic. Interview techniques such as Applied Cognitive Task Analysis are intended to make explicit expert knowledge that has become automatic. A fireman is likely to be proud of his knowledge and pleased when the interviewer can identify some piece of knowledge he had forgotten was special. Using ACTA to interview KT about her "expertise" (which it is) in the use of language however, KT thinks of her knowledge as just common-sense. The knowledge was implicit knowledge – a set of learnt skills as to how to converse. What we were after was exactly that common-sense knowledge in an explicit form so we could model and use it. Unfortunately it is common sense also in that is common to all – it is knowledge that is shared and KT knows that. Interviewing people about their common-sense knowledge, they quickly become suspicious about the interviewer's motivations. Why is he asking such "dumb" questions?

The lessons from this were that it was precisely the implicit social skills in conducting a conversation that were important but also difficult to get at in an explicit form. Just as one can be able to ride a bicycle but not know how one does it, one can conduct a sensible social interaction whilst not being able to specify how one does this. The very ubiquity of this skill hides its subtlety and complexity.

2.2 Ethnomethods

The CA4NLP project applied an ethnomethodological variant of Conversation Analysis [11] to analysing records of conversations such as those produced by the KT experiments. This approach is predicated upon the notion that the researcher is a "member of the same community of practice" as the discussants, and hence has access to the import of their utterances. Thus, for example, a researcher's introspections about whether or not some communicative act of KT's was polite is valid evidence because both get their knowledge about the purpose of communicative acts from the same common pool. I do not need to ask KT about her internal reasoning because it is the external effect that matters and I have direct access to its significance. KT is right: I could give as good an answer to my own dumb questions as she.

This method also implies a shift from a mechanistic view to a functional view. When it comes to engineering spoken language interfaces, rather than trying to access the internal reasoning of the speaker as the KT experiments attempted, we want to look at and model the way a social agent engages with the community of practice in which it operates. Although engineering more as a process of adaption of function than design will make some engineers uncomfortable, this is common practice for long-standing artefacts that inhabit complex niches, such as sailing yatchs – nobody designs a yatch from first-principles but rather adapts and tunes existing designs, tinkering with each aspect in turn. What matters is how the yatch functions within the complex environment of winds and water. The same applies to computers that act in our social space. A computer that says "no records match your request" might be being informative [12] but is it playing by the rules of social engagement? Using the terminology from Conversation Analysis, what is the *work done* by "no records match your request" and is it all and only what the expression was designed to do in the current context?

The methodology of Conversation Analysis is for the scientist to capture naturally occurring text or speech of interest and ask "Why this, in this way, right here?" Whilst using introspection as a means to assess scientific truth is a bad idea, introspection about community knowledge is fine and provides detailed descriptions of the function of utterances in context. Thus the CA4NLP project illustrated the use of introspection to leverage understanding about utterances. It marks a shift away from attempting to access an internal or foundational model, but rather capitalises upon the function of utterances in context. It is the function of utterances that is constrained by common usage, not the cognitive processes that give rise to them.

The trouble with Conversation Analysis however is exactly its strength in that it provides a valid means of studying anything and everything. It does not provide any guidance on what is *critical* to the structure of a conversation.

2.3 HCI and Grounded Theory

The SERA project put a talking "rabbit" in older people's homes and collected video of the resulting real human-robot interactions. 300 or so recordings of people interacting with their rabbits were collected. The experiment had three iterations of: placing the rabbit, recording the interactions, assessing the success of the system, and improving the system software based on the assessment.

The motivation for the project was to see how different research groups would go about this process. In general, all the groups could find interesting things to write about the data, but the process of improving the system was primarily driven by those with an HCI background who would, in the tradition of design-based engineering, simply have an idea that could be tried. This creative process often worked, and would be followed by a quantitative evaluation, but felt quite unsatisfactory when it came to understanding what is going on.

The understanding that did feel like progress actually came from qualitative methods such as Grounded Theory [13] in the form of detailed analyses of how particular conversations unfolded in those contexts. In particular people are very good at noticing when a conversation is NOT right. As an expert I can tell you that I wouldn't say "no records match your request" in a given context and it is this data that needs to be the raw material on which we base a science of machines in social spaces. However, this micro-level of detail poses a problem when one needs to utilise the knowledge, for example in terms of suggesting improvements to CA scripts. The detail needs to somehow be accumulated in a more comprehensive social ability.

In some preliminary experiments in the SERA project, people were asked to say what happened in a video recording we had of people interacting with one of the SERA rabbits. This initially did not work very well because, although the plot in a film or play is easily identified and summarised, natural recordings are just not that interesting and rather messy. Instead recordings where things go wrong was chosen. This made the 'crux' of the story more salient.

2.4 Summary of threads

From the above experience we draw out several lessons. We see the importance of the shared culture in terms of the common folk theory about what is happening, however we also see that this common knowledge is implicit and not very accessible via direct interrogation. We see the importance of examples learnt in context, in particular in terms of their functional fit to the social circumstances. Finally this suggests that, in order to transfer this implicit knowledge we might have to mimic the learning that usually happens within the social sphere in terms of making mistakes and repairing them.

In order to make better conversational agents they will have to be inducted into the society they are going to inhabit. Clearly, in general, this is extremely hard and takes humans a couple of decades of time but here we might be aiming for an agent that copes tolerably well (on the level of a polite 6 year old) in a single context (or a very restricted range of contexts). Here we aim to imitate the cycle of trial, error and repair on a small scale, hoping to make up for the small number of cycles with a more intelligent repair stage composed of analysis with repair leveraging some of our own innate understanding of social behaviour. Each iteration in a particular context will (on the whole) result in an incremental improvement in social behaviour. The hard part of this cycle (other than the number of times it may have to be done) is the analysis and repair stage. We will thus concentrate on this in this section.

3 Capturing the implicit

The idea presented in Figure 1 is to iteratively improve an *in situ* CA, each iteration through allowing a bit more of the explicit *and* implicit knowledge concerning the appropriate social behaviour to be captured in the knowledge base and hence used to tune the CA rules. Each iteration the CA, in its current state of development, will deployed and new records of its conversation with humans made, since it is difficult to predict the full social effect of any change. This iterative cycle imitates, in a rough manner, the way humans learn appropriate social behaviour: observing others, noticing social mistakes and iteratively adapting their behaviour.

Clearly there are several parts of this cycle that could be discussed in detail. However, here we will concentrate on motivating and outlining how the user-interface that prompts and structures the review of the conversational records by the native expert. The nub of this process is how to elicit the, largely implicit, knowledge about social behaviour using the responses of the third party reading and reacting to the records of the conversation.

3.1 Vygotski

Vygotski's insight used here is that plays and novels exist because they provide plausible accounts of human behaviour. Theatre is the flight-simulator of life [14] and provides a means of exercising our ability to understand the motivations and behaviour of others. We do think about other minds when we communicate – indeed it turns out to be a critical skill [15–17] – and we do it in terms of beliefs, desires and other mental

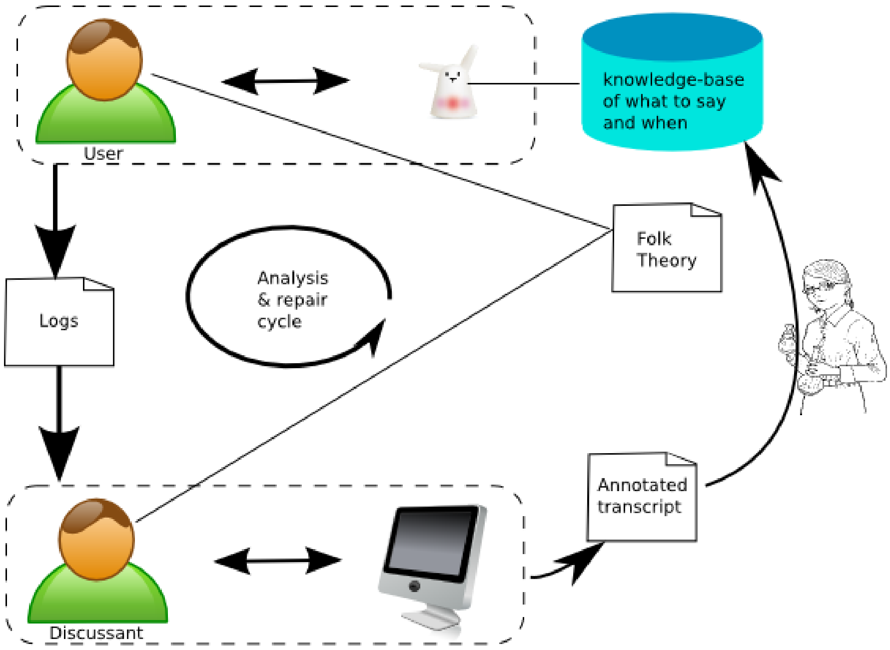

Fig. 1. Summary flow chart of the proposed iterative method

attitudes. What is more, we *expect our conversational partners to do the same*, with the same model. When it comes to communication, the truth of our folk model of other people's thinking doesn't matter; what matters is that it is shared. Rather than looking inside KT's head to see how she would deal with social relations, the idea is to look at some kind of collective understanding of events – what is the shared knowledge that creates the context against which a human social actor figures out the significance of communicative acts? Rather than sitting in an arm-chair and classifying utterances according to the effect they have on a idealised conversational partner, the idea is to look at real interaction data and document the effect in context. Rather than classifying utterances as $REQUEST\_INFORMATION$, or $GREETING$ [18], the idea is to record the "work done" by utterances in the place they are produced. This can be done by any member of the community of communicators and does not require a scientific theory. Consider this example of a conversation between a doctor and a patient taken from the Conversation Analysis literature:

```
Patient: So, this treatment; it won't have any effect on
us having kids will it?
Doctor: [silence]
Patient: It will?
Doctor: I'm afraid the...
```

The “work done” by the silence is of course to disagree and some might be tempted to mark it up as an explicit answer, but there are many different things that the doctor could say at this point, with a wide range of “semantics” but all with the same effect.

The Vygotski argument is that human story telling gives sufficient detail of events that any socialised human (who is part of the same community) can fill in the gaps to produce a set of linked causal relationships for the story to make sense. This requires contextual knowledge (e.g. teddy bears are toys and children like to play with toys) and ”hard-wired” knowledge (e.g. children often *want* things and *act* in ways that bring them about).

One might think that human-machine interactions would be less fraught and thus simpler. Indeed those working on commercial spoken language interface design try very hard to make this true using techniques such as menu choice. However, even in the DARPA Communicator data where the systems were only slightly more natural than those one might find in a bank, there are examples where the work done by an utterance such as “no” goes well beyond what might be seen as the semantics of the utterance [19]. One can not escape the importance of social etiquette.

It is this process, and the interface to support it, that we will now describe.

3.2 An example

Consider some video data captured spontaneously (Figure 2) during the development of the SERA set-up.

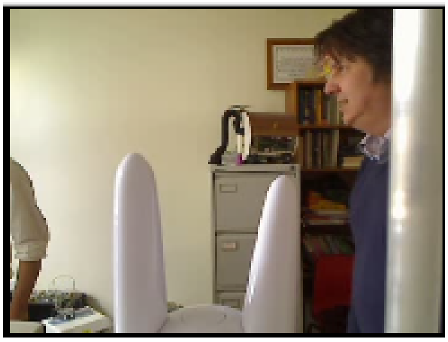

Fig. 2. Mike and the rabbit talking with Peter

To set the context, “Peter and Mike have been talking in Peter’s office where he has a robot rabbit that talks to you and that you can talk to using picture cards.” Two narrators were given this sentence and asked to watch the video. They were then asked to, independently, finish the story in around 200 words. The resulting stories appear in Figure 3.

There are many differences, and many things were left out entirely. There does however appear to be general agreement on core events. Neither narrator mentioned

Narrator 1	Narrator 2
It is time to go home so Peter takes his keys from the rabbit. Mike notices this and says “Isn’t it supposed to say hello?” Peter is about to say something when the rabbit says: “Hello, are you going out?” Peter replies that he is (using the card and verbally) and the rabbit tells him to have a good time, bye. Mike picks up a card and shows it to the rabbit, but nothing happens. He thinks this make sense as the rabbit has said goodbye but Peter thinks it should work and shows the rabbit another card. Mike sees that he has been showing the cards to the wrong part of the rabbit and gives it another go. Still nothing happens and Mike tries to wake it up with an exaggerated “HELLO!”. Peter stops packing his bag and pays attention. Mike tries getting the rabbits attention by waving his hand at it. Still nothing happens. Mike looks enquiringly at Peter as if to ask “what’s happening” He says “that’s a new one” and goes back to his packing. Mike takes his leave at this point. Peter finishes his packing, and, as he leaves says to the rabbit “You’re looking quite broken.”	Peter is about to do something to wake the rabbit up again and as he is about to speak, it says hello. Peter gestures to Mike that it is now talking as expected. Peter presses the video button to record the interaction. Mike laughs as it talks. It asks Peter if he is going out, to which he responds verbally that he is, showing the rabbit the card meaning yes. Seeing Peter’s interaction, Mike tries using the cards to interact with the rabbit himself. It does not respond and Mike suggests that this is because it has said goodbye and finished the conversation. Peter tries to reawaken the rabbit with another card. Mike sees that he had put the card in the wrong place. He tries again with a card, after joking that the face card means “I am drunk”. Peter laughs. When the rabbit does not respond, Mike says “hello” loudly up to the camera. Peter says he is not sure why there is no response while Mike tries to get a reaction moving his hand in front of the system. They wait to see if anything happens, Mike looking between the rabbit and Peter. When nothing happens, Peter changes topic and they both start to walk away. Mike leaves. As Peter collects some things together, walking past the rabbit, he looks at it. Before leaving the room he says to the rabbit “you’re looking quite broken”.

Fig. 3. Two narrative descriptions of the same event.

1.	Peter is about to say something and is interrupted by the rabbit
2.	the rabbit asks if he is going out, Peter’s verbal and card response
3.	the rabbit says bye
4.	Mike’s attempt to use a card and the non-response of the rabbit
5.	Mike’s explanation (that the rabbit has already said bye)
6.	and Peter showing the rabbit another card
7.	Mike sees that he has been showing the card to the wrong part of the rabbit and has another go
8.	the rabbit does not respond
9.	Mike says “Hello” loudly
10.	Peter acknowledges it doesn’t look right
11.	Mike tries again by waving his hand in front of the rabbit
12.	no response from the rabbit
13.	Mike looks at Peter
14.	They give up
15.	Mike leaves
16.	Peter leaves saying “You’re looking quite broken” to the rabbit

Fig. 4. The third-party common ground.

the filing cabinet nor the clothes participants were wearing. No comment on accents or word usage; no comment on grammatical structure nor grounding, nor forward and backward looking function. Whatever it is that the narrators attend to, it is different to the type of thing that appear in classic annotation schemes. It does however seem to be shared and, the claim is, shared by the community of practice at large. Both the narrators and the participants are working from a shared theoretical framework – not from raw undifferentiated data – that guides and selects which sense-data is attended to. However this shared framework is implicit.

Accounts of the action in the video data as written down by the narrators are of course *descriptive* in that they are written to 'fit' past events. The claim is that they are also predictive. If Mike wants to use the system, then it would be surprising if others did not want to. If failure to work causes disappointment in Mike, it is likely to also cause it in others. Having a predictive model of events we are well on the way to having *prescriptive* rules that can be used to drive conversational behaviour.

But first however let's look at how we might move more formally from the stories in Figure 3 to the summary in Figure 4.

3.3 An interface to support capture of social knowledge

The problem is that even if they observe the same thing, they may not describe it in the same way and, unless the descriptions are the same, a machine cannot recognise them as the same. In the example above the two observers produced two narrative descriptions and it is claimed they are the same, but how would one *measure* the sameness? Without a machine that can understand what is written, human judgement is involved and claims of researcher bias are possible. How might comparative narratives [20] be produced that are the same to the exacting standards required for machine understanding?

The proposal, should one want to re-do this preliminary experiment properly, is to use the techniques seen in industrial machine translation for the production of operator and repair manuals. Companies like Mitsubishi and Caterpillar [21] have systems that allow them to produce manuals in one language and then, at the push of a button, produce the same manual in all of the languages for countries to which they export. The way this is done is to have the author of the manual write in a restricted version of the source language and provide the tools to guide the writing process. The process of authoring with such tools will be familiar to us all because modern text editors provide spelling and grammar checking assistance in much the same way. The primary differences being of course that the list of recognised words is much smaller and the grammar rules much stricter, and the process of breaking those rules is not simply for the system to ignore it, but to ask the user to add the new word or expression to the system. For instance the author might really want to use the term "airator" and the system would allow that but ask the author if it is a noun or an adjective, a count noun, what its semantic preferences are, and if it is masculine or feminine in French. The word would be added to the lexicon and, the next time an author wanted to use it, the system would have enough detail to translate it correctly or ask this new author how it should be used in the current context.

If one wanted to re-do the experiment above more formally, the approach would be to reproduce a "translators work bench" and, rather than having it translate to another language, have it "translate" to a different style in the same language. This authoring process works for machine understanding for translation; there is no reason to think it wouldn't work for this new application if one really wanted to do it. But why bother? The ultimate aim is to script dialogue for synthetic characters and the proposal is that, rather than stopping at narrative descriptions, the system would go on to explore counterfactuals.

3.4 Narrative descriptions capturing context

The aim is to classify utterances as the same in context and hence be able to program an agent to give a particular response to any input from the same class. Using a classic annotation scheme one might decide that if its conversational partner produces something in the class of $QUESTION$, then the agent should produce an $ANSWER$. This functionalist model of sameness applies to everything from chatbots in which something like regular expressions are used to recognise inputs are the same, through to full planning systems such as TRAINS [22] in which input recognition is set against the current goals of the system. The variation proposed here is that the functionalist definition of sameness is embedded in narrative. Two expressions are the same if and only if, for every narrative in which expression #1 occurs the outcome of the story would not change if expression #2 was used.

Given such a definition of sameness, it is only in trivial cases that expressions will be universally the same. It is far more likely that expression #1 and #2 will be equivalent for some narratives and not others – the equivalence is context dependent, and this provides an opportunity to question an observer about the features of the context that determine when an existing response to input might or might not be appropriate for another input.

As an example of the type of thing we have in mind Figure refCTAprobes gives a table showing the type of question that was asked of KT. It would appear that some of these questions would be a useful way to explore context with our observers and, importantly, the questioning could be automated. An observer might provide a narrative description of a particular recording of an interaction and, at some point in that description the computer might say S where the rules being used by the machine might have equally produced S'. An annotator's work bench could ask the human if S and S' would be functionally equivalent in the narrative given. If not, the workbench could ask what (in the context) makes S' inappropriate, and perhaps ask the annotator to develop a rule that distinguishes the context for S and S'. Similarly the system could ask the observer if he or she can formulate an alternative to S and S' that would be better, and develop a rule to distinguish the alternative utterance from S and S'.

The above gives a flavour for the proposed work bench designed to enable non scientists to use their expert knowledge of language use to create context dependent rules so the system can decide what to say when. The aim is to combine the direct contact with the data normally seen in an annotation tool such as Anvil [23] with the creative process of scripting conversation for the agent. In effect the aim is to formalise the process (and add some theory) that people use when they script chat-bots using AMIL by pouring over log files.

Fig. 5. O'Hare et al 1998 - the revised CDM probes.

Goal specification	What were your specific goals at the various decision points?
Cue identification	What features were you looking at when you formulated your decision?
Expectancy	Where you expecting to make this type of decision during the course of the event? Describe how this affected your decision-making process
Conceptual model	Are there any situations in which your decision would have turned out differently? Describe the nature of these situations and the characteristics that would have changed the outcome of your decision.
Influence of uncertainty	At any stage, wee you uncertain about either the reliability or the relevance of the information that you had available? At any stage, were you uncertain about the appropriateness of the decision?
Information integration	What was the most important piece of information that you used to formulate the decision?
Situation awareness	What information did you have available to you at the time of the decision? What information did you have available to you when formulating the decision?
Situation assessment	Did you use all the information available to you when formulating the decision? Was there any additional information that you might have used to assist in the formulation of the decision?
Options	Were there any other alternatives available to you other than the decision that you made? Why were these alternatives considered inappropriate?
Decision blocking - stress	Was there any stage during the decision-making process in which you found it difficult to process and integrate the information available? Describe precisely the nature of this situation.
Basis of choice	Do you think that you could develop a rule, based on your experience, which could assist another person to make the same decision successfully? Why/Why not?
Analogy/generalization	Were you at any time, reminded of previous experiences in which a *similar* decision was made? Were you at any time, reminded of previous experiences in which a *different* decision was made?

4 Conclusion – Towards the Iterative Embedding of Implicit Social Knowledge

This proposed approach seeks to take seriously the subtlety of social behaviour, resulting from the "double hermeneutic" which relies on the fact that encultured actors will have a ready framework of how to interpret the social behaviour of others, including the expectations that others will have of them. In particular it is important how it is that social knowledge is embedded within a complex of social relations and knowledge, which makes it hard to formalise *in general*. We do not expect that this will be easily captured in a "one-off" analysis but require a iterative approach based on repair. The difficulty of the task means that a number of approaches will need to be tried to leverage little bits of social knowledge each iteration. The key parts of this are the interactive capture of social information from a third party and the use of that knowledge to inform an update of the CA rules. We have not talked about the latter here – currently it will require significant programming skill. The ultimate aim would be to eliminate this programmer, so that this iterative process could be used by non-experts, utilising their own implicit expertise, to socially "educate" their own CA. This is illustrated in Figure 6.

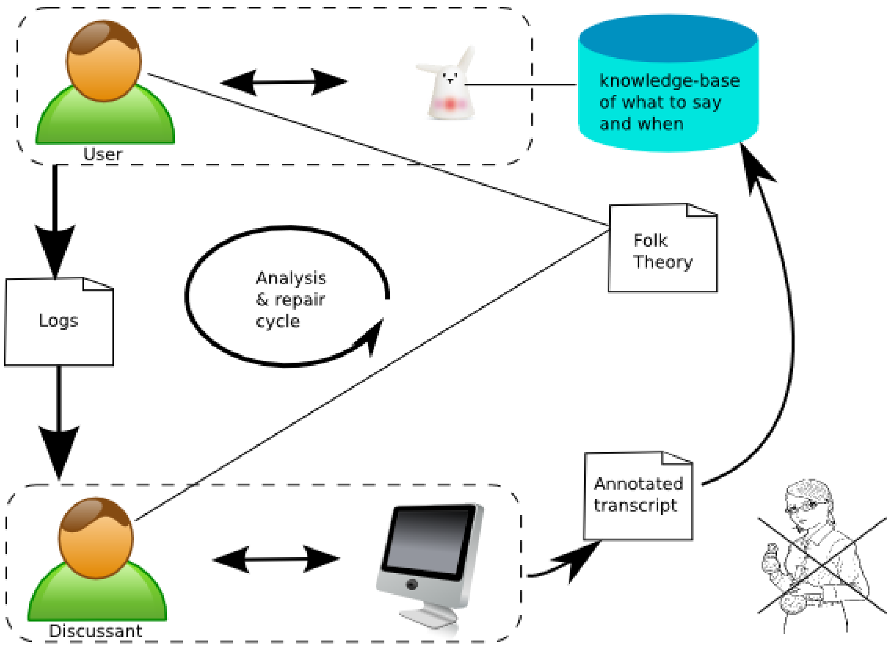

Fig. 6. Flow chart of the process without the programming expert

References

1. de Angeli, A.: Stupid computer! abuse and social identity. In de Angeli, A., Brahnam, S., Wallis, P., eds.: Abuse: the darker side of Human-Computer Interaction (INTERACT '05), Rome (September 2005) http://www.agentabuse.org/.
2. Mel'cuk, I.: Meaning-text models: a recent trend in soviet linguistics. Annual Review of Anthropology **10** (1981) 27–62
3. Hovy, E.: Injecting linguistics into nlp by annotation (July 2010) Invited talk, ACL Workshop 6, NLP and Linguistics: Finding the Common Ground.
4. Young, S.J.: Spoken dialogue management using partially observable markov decision processes (2007) EPSRC Reference: EP/F013930/1.
5. Edmonds, B.: Complexity and scientific modelling. Foundations of Science **5** (2000) 379–390
6. Edmonds, B., Cershenson, C.: Learning, social intelligence and the turing test: Why an 'out-of-the-box' turing machine will not pass the turing test. In Cooper, S.B., Dawar, A., Lwe, B., eds.: Computers in Education. Springer (2012) 183–193 LNCS 7318.
7. Edmonds, B., Bryson, J.: The insufficiency of formal design methods – the necessity of an experimental approach for the understanding and control of complex mas. In: Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'04). ACM Press, New York (July 2004) 938–945
8. Wallis, P., Mitchard, H., O'Dea, D., Das, J.: Dialogue modelling for a conversational agent. In Stumptner, M., Corbett, D., Brooks, M., eds.: AI2001: Advances in Artificial Intelligence, 14th Australian Joint Conference on Artificial Intelligence, Adelaide, Australia, Springer (LNAI 2256) (2001)
9. Militello, L.G., Hutton, R.J.: Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. Ergonomics **41**(11) (November 1998) 1618–1641
10. Brown, P., Levinson, S.C.: Politeness: Some Universals in Language Usage. Cambridge University Press (1987)
11. ten Have, P.: Doing Conversation Analysis: A Practical Guide (Introducing Qualitative Methods). SAGE Publications (1999)
12. Traum, D., Bos, J., Cooper, R., Larson, S., Lewin, I., Matheson, C., Poesio, M.: A model of dialogue moves and information state revision. Technical Report D2.1, Human Communication Research Centre, Edinbrough University (1999)
13. Urquhart, C., Lehmann, H., Myers, M.: Putting the theory back into grounded theory: Guidelines for grounded theory studies in information systems. Information Systems Journal **20**(4) (2010) 357–381
14. Bennett, C.: But mr darcy, shouldn't we be taking precautions? Keith Oatley (quoted in) (July 2011)
15. Grosz, B., Sidner, C.: Attention, intention, and the structure of discourse. Computational Linguistics **12**(3) (1986) 175–204
16. Eggins, S., Slade, D.: Analysing Casual Conversation. Cassell, Wellington House, 125 Strand, London (1997)
17. Tomasello, M.: Origins of Human Communication. The MIT Press, Cambridge, Massachusetts (2008)
18. Jurafsky, D., Shriberg, E., Biasca, D.: Switchboard swbd-damsl shallow- discourse-function annotation coders manual. Technical Report 97-01, University of Colorado Institute of Cognitive Science, Colorado (1997)
19. Wallis, P.: Revisiting the DARPA communicator data using Conversation Analysis. Interaction Studies **9**(3) (October 2008)

20. Abell, P.: Comparing case studies: an introduction to comparative narratives. Technical Report CEPDP 103, Centre for Economic Performance, London School of Economics and Political Science, London, UK (1992)
21. Kamprath, C., Adolphson, E., Mitamura, T., Nyberg, E.: Controlled language for multilingual document production: Experience with caterpillar technical english. Proceedings of the Second International Workshop on Controlled Language Applications **146** (1998)
22. Allen, J.F., Schubert, L.K., Ferguson, G., Heeman, P., Hwang, C.H., Kato, T., Light, M., Martin, N.G., Miller, B.W., Poesio, M., Traum, D.R.: The TRAINS project: A case study in defining a conversational planning agent. Journal of Experimental and Theoretical AI **7**(7) (1995) 7–48
23. Kipp, M.: Gesture Generation by Imitation - From Human Behavior to Computer Character Animation. PhD thesis, Saarland University, Saarbruecken, Germany (2004)