

Conversation in Context: what should a robot companion say?

Peter Wallis, Viktoria Maier

Dept of Computer Science,
The University of Sheffield,
Sheffield, S1 4DP, UK

firstInitial.surname@dcs.shef.ac.uk

Sarah Creer and Stuart Cuningham

ScHARR (Health Services Research),
The University of Sheffield,
Sheffield, S1 4DP, UK

firstInitial.surname@sheffield.ac.uk

Abstract

Language as used by humans is a truly amazing thing with multiple roles in our lives. Academics have tended to focus on the way languages convey meaning, and disciplines that come new to the problem such as computer science tend to start with reference semantics and progress to models of meaning that look mathematical and hence solidly academic. Language as used is however beautifully messy. People sing, they lie and swear, they use metaphor and poetry, play word games and talk to themselves. Is there a better way to look at language? Interdisciplinary research is hard not only because each discipline has its own terminology, but also because they usually have different interests. Those of us interested in spoken language interfaces (computer science) however have a shared interest with applied linguistics in how language works in situ. This paper outlines a theory about how language works from applied linguistics and shows how the theory can be used to guide the design of a robot companion.

1 Introduction

In 2005 and 2006 some of us were involved in a workshop series on computers and abuse. Our motivating interest was in why people swear at chat-bots. This is not some minor fluctuation - de Angeli looked at transcripts from Jabberwacky and found 20% of the *words* were abusive [de Angeli, 2005]. Indeed it seems this abuse is not species specific. In experiments with an Aibo and dogs, there is dramatic footage of the dog throwing the machine across the room. It seems that we animals do not like machines. What is happening? In one of the workshops it was proposed that the abuse we observe might be part of some unconscious mechanism that enables intra species cooperation [Wallis, 2005]. In the same way as termites unthinkingly (presumably) follow rules that result in large complex artifacts, might there be simple rules of human behaviour that in some way enable our more grand achievements.

It turned out that a description of such rules can be found in applied linguistics.

Conversation Analysis (CA) is a methodology with a strong commitment to naturally occurring data and the ethnomethodological variants have strong links with anthropology. Its techniques are aimed at noticing and explaining the everyday - the things that we do without thinking. CA is usually associated with the very low level details of conversation - the nature of turn taking, the structure of openings, the notion of adjacency pairs and so on - but Seedhouse [Seedhouse, 2004] sums up the finding of CA over the years with the macro observation that an utterance in a conversation either “goes seen but unnoticed, noticed and accounted for, or risks sanction.” In the case of chat-bots, this sanction takes the form of swearing.

The theory is that language works in the first instance in much the same way as we computer scientists think with words referencing things and action in the world. In language in use, your conversational partner (CP) often produces an utterance that goes **seen but unnoticed**. He or she will provide answers to questions, will return greetings and give explanations as you expect. When that fails - when your CP’s response is in some way out of the ordinary - you, the listener, notice the utterance and work very hard to find an explanation for what he or she said. An utterance becomes **noticed and accounted for**. It is worth quoting at length the Eggins and Slade [Eggins & Slade, 1997] description of just how much emphasis we humans place on sequential relevance:

Consider, for example, the two turns at talk below:

A What’s that floating in the wine?

B There aren’t any other solutions.

You will try very hard to find a way of interpreting B’s turn as somehow an answer to A’s question, even though there is no obvious link between them, apart from their appearance in sequence. Perhaps you will have decided that B took a common solution to a resistant wine cork and poked it through into the bottle, and it was floating in the wine. Whatever explanation you came up with, it is unlikely that you looked at the example and simply said “it doesn’t make sense”, so

strong is the implication that adjacent turns relate to each other.

This is where, in computer science terms, things get untidy and where, for instance, the need for context is found. It is interesting to note that the need for context is in this second pass at understanding.

Finally, if you fail to account for your CP's utterance, then your CP **risks sanction**. When your CP is a chat-bot, that sanction is likely, it seems, to take the form of swearing.

2 Providing an accounting for

This notion that each conversational partner's utterance is either seen but unnoticed, noticed and accounted for, or risks sanction, provides an explanation for the considerable interest in the literature in having conversational characters that are *consistent*. It has been claimed that clarity and consistency are among the most important features in an agent (Norman, 1988; Tognazzini, 1992; Winograd, 1996) and allows people to predict what will happen when they engage with the system and lightens cognitive load (Fiske & Taylor, 1991). Isbister & Nass (2000) state that it becomes easier to deal with an agent, if the agent is consistent. They point out that consistency also means consistency with a user's expectations (Tognazzini, 1992, pp. 250, 251). Thus, good agent design should work with the already existing set of expectations that a person brings to his/her interaction with the object, creating intuitive "affordances" (Norman, 1988). Nass & Brave (2005) point out that inconsistency leads to difficulties in perception and processing. These claims are intuitively sensible, but the realisation that people work hard to account for an agent's action provides the mechanism by which consistency helps. A consistent character is one that provides an accounting for its behaviour.

This notion of seen but unnoticed, noticed and accounted for, or risks sanction, also provides a useful framework for analysing system performance. As a classic example of a machine's interaction eliciting sanctions, here is a conversation from the 1996 Loebner Competition:

PROGRAM G'day mate.
JUDGE G'day. How's it going?
PROGRAM My mother once said "How it is going is a mystery." She was a bit nuts.
JUDGE Uh-huh
PROGRAM Are you still learning to speak, my friend?
JUDGE NO, no. But your anecdote took me a bit by surprise.
JUDGE How was your morning?
PROGRAM Oh well, that's a pity!
JUDGE I'm sorry; what's a pity?
PROGRAM Apologies accepted.
JUDGE I feel compelled to insult you and we've only been talking a minute.

PROGRAM We are bloody well talking, aren't we?

JUDGE Oh baby. We sure are. So here are some keywords for you: shit mother baseball.

...

This was the winning entry from that year and of course there are more recent and better funded projects. The DARPA Communicator programme was a brilliant success in that it reflected just how much we do not know about language in use. The participating organisations had access to world experts on spoken language interfaces, and effectively unlimited resources, but failed to get anywhere near the vision as set out by the programme coordinators. The observation is that these systems did mostly well when the human stuck to answering the system's questions and the automatic speech recognition (ASR) worked. That is, we know how to handle the seen but unnoticed. When the grounded knowledge was not as the script designers expected however, the system did not have the world knowledge to understand the user's move to repair the situation and the system would produce a response that the user could not account for. The result was sanction. In Communicator, swearing at the system is rare (surprisingly) for reasons that can only be guessed at. One can however see (hear) sanction happening in the recordings with either the user expressing exasperation or simply hanging up the phone, but it is also reflected in the post call survey in which a significant proportion of users did not want to "use the system on a regular basis." "not wanting to use the system on a regular basis" [Wallis, 2008].

As someone with connections with the dialog systems community it is tempting to blame the speech recognition technology but there is a growing body of evidence that word error rates are not the problem [Wallis *et al.*, 2001; Skantze, 2007; ?]. The problem is not "trouble in text" such as failed ASR or a lack of world knowledge, but rather people get annoyed with spoken language interfaces because the systems we produce fail to repair the trouble. Trouble happens all the time in human-human communication but this is not annoying in itself as long as one's conversational partner is observably trying to help. That is, as long as your CP's utterances can be accounted for, you don't even notice that trouble occurred.

The mechanism for accounting for can be both tactical and strategic. Eliza and Parry were very successful in that user satisfaction was very high compared to modern day systems. The mechanism was strategic in those systems in that they provide an accounting for their behaviour – in the first case because the role of psychologist accounts for the endless stream of personal questions, and in the second because being paranoid accounts for the system's odd responses and interests. At the tactical level, at least one of the systems in the Communicator programme would, if the conversation got recognisably confused, say the "network connection seems to be down" and ask the user to try again later, effectively accounting for its failure to provide a flight booking. Emotion too can play a

tactical role. Eugene was a virtual cuttlefish that expressed emotion through colour. At the strategic level, Eugene’s “personality” was rather vain and he liked to talk about his pretty colours. Stuck for something to say Eugene would ask you what you thought of his colours. This otherwise rather strange conversational gambit worked because his persona accounted for the behaviour and so it did not seem strange at all. If you were rude to Eugene, he would change the subject giving the impression that he was offended and hence accounting for the change of subject. Once again, the accounting for is so effective that to us humans (as opposed to us researchers) it is hard to notice what the issue is.

It seems our natural language interfaces can get away with many things including non understanding and poor ASR if they are set up so that the user can account for their behaviour. In our tool box of techniques are persona, role and emotion, but in human-human dialog we also find explicit discussions to account for possibly unexpected behaviour. In a set of experiments based around booking cars out of the Division’s car pool [Wallis *et al.* , 2001], there were explicit discussions to account for delay (often waiting for the computer to do something) and indeed about high level plans. Our subject would describe to the caller how she was going to handle their call, followed by “is that ok?” - which was not really a question but was an opportunity to provide more information that would change her plan. One might have said “No, just do ... ”, but such an unaccounted for move would certainly risk sanction.

Finally, although someone will work very hard to find an interpretation of their conversational partner’s utterance, this requires that the person is committed to the conversation. I will work hard to figure out why a CP said what he or she said, in the way it was said, but to do that requires that I know that I am being addressed. This is not an issue for a chat-bot on a website or a system set up for experiments in a laboratory, but becomes a significant issue for an interactive artifact that is permanently in someone’s home.

3 SERA

The SERA project is an EU programme funded through the FP7 ICT call 3 on cognition and robotics to look at “social engagement with robots and agents.” The focus has turned out to be on robots rather than embodied conversational agents and the robot of choice is a Nabaztag. The Nabaztag is a commercially produced talking head from Violet in the style of kismet, the iCat, or indeed Foust’s talking skull. It is a stylised rabbit with expressive ears and a set of multi colour LEDs (see Figure 1) and is marketed as the world’s first internet enabled talking rabbit. The rabbit connects to the Violet server via a wireless router and can run several applications including receiving SMS messages, weather reports, tai chi, and streaming selected radio or blog sites.

Our aim is to study long term relationships between



Figure 1: The Nabaztag on a pedestal (with array microphone, web-cam, and a PIR sensor)

people and robot “companions” and the intention was to put a Nabaztag in an older person’s home and see what happens. This is not as straight-forward as it may first appear as older people are not the intended market niche for this product. Mival *et al.* find that products such as the Furby and Aibo tend to end up “in the back of the cupboard with the batteries out,” and our aim was to study the way relationships with a machine develop over time. Mival *et al.* suggest that a synthetic companion needs to have a perceived use - even if that use is not really useful. Unless one likes technology for its own sake, the user needs a reason for having the thing on.

The study takes the form of three iterations of field studies. At each stage the system will be deployed in users’ houses and remain with them for ten days. The observations made after each iteration will inform the system development for the following iteration.

4 Designing the interaction

The critical thing is that the user of our technology can **account for** the system’s behaviour when the unexpected happens and he or she notices. The features that may provide an account include role, persona, emotion, theory of mind (“Oh it doesn’t know that ...”) and explicit explanation.

The use of the robotic agent in this case is to provide an activity and exercise application. Following the British Heart Foundation literature [Lewin *et al.* , 2005]. designed for individuals recovering from heart failure and wishing to increase their activity levels, the participants were provided with interactions which monitored an activity plan. The rabbit was also de-

signed to provide short small talk type interactions, provide a weather report and pass on messages from the researchers. Evidence from the data collected in the first iteration shows that, specifically, the activity-related interactions did not promote engagement of the users with the device.

Evidence from the post-trial interviews support the view that the users did not think that this process was relevant to them as they were already committed to following their own activity plan.

P1: “But I never forget that I am going to Aqua [aerobics], because there is only so many sessions a week.”

P2: [the rabbit] “asked me if I have had done my activity, which I do if I want to and I very rarely don’t. And if I don’t do there is a very good reason. I didn’t need to be told.”

One explanation for this lack of perceived usefulness takes into account the transtheoretical model (TTM) of behaviour change [Prochaska & Velicer, 1997]. The TTM consists of five stages of change: precontemplation, contemplation, preparation, action and maintenance. The activity plan task as designed for the first iteration of the SERA project clearly fits in with the preparation and action stages of the model. By responding to the recruitment advert, the individual has moved beyond the stages of precontemplation and contemplation. The evidence from the data collection suggests the participants in the first iteration are in the maintenance phase of this model; they have a set activity plan which they are currently following and are maintaining that level of activity. When asked the question “Did you stick to the amount of activity that was in your activity plan?”, there were zero occurrences of a ‘no’ response in the video data. The users are already leading what they consider to be a healthy active lifestyle, do not see themselves within the group of individuals at whom this device is aimed and are participating in the project to help others via this research.

P1: “the whole point of this is to see if, you know, us old ‘uns just sit on the settee all day long doing nothing isn’t it? ‘Cause some of them do.”

P2: “I do what I do and more ... I will walk into town, go swimming. And then I rush around the rest of the time.”

The task for the first iteration was not consistent with the expectations or requirements of the user and therefore not useful or relevant to them. To account for this and also to include those participants who are not yet recruited for the second iteration of the data collection stage, the application should be useful for and consistent with the expectations of people who are in any of the preparation, action or maintenance stages.

The concept of ‘regression’, the move backwards to a previous stage, is relevant at all stages of the model and the prevention of this could be the focus for the new task. A vital construct in the prevention of regression is that of ‘self-efficacy’ [Bandura, 1977], the

context-specific confidence that an individual has in their ability to fulfil their goals. Maintaining a high level of self-efficacy will help to keep the individual in the current stage. It is therefore hypothesised that the activity related interactions should effectively provide confidence boosting or confidence maintaining behaviour. There are many occasions in the video data and interviews where the individuals were keen to talk about how much exercise they did and tell the rabbit more about their activity. Providing a way for this information to be input to the rabbit and feeding that information back in a summary to the participant will remind them of how much they have achieved which will more accurately match the users’ image of themselves and their current stage in the model. It is hypothesised that this will make the information relevant to them and increase the perceived usefulness and therefore engagement with the robotic agent.

5 Role and persona

The iterative process involved in the SERA project allows the observations of recorded interactions with the robot to inform the next stage of design to increase the potential for social engagement. Ideally our robot companion would have a fully fleshed out human character with a history, likes and dislikes, friends and relations, and be able to talk on any subject from quantum physics to the British soap opera Coronation Street. Considering the technical implications of the ideal and the need for the application to work reliably for it to be acceptable to the user, it is necessary to limit the scope in a way that the human can account for.

A popular mechanism to limit the scope is to have the system provide a service and take on a role as a service provider. For instance, phoning a travel agent to book a flight, one can account for the person on the other end of the phone not wanting to talk about Coronation Street. For the robot companion, this notion of a constrained role is harder as the system is in the user’s space permanently.

For iteration 1, the system deployed used yes/no and a video button. Specifically, the agent’s role was to help the user to maintain the level of planned exercise, by passing on information from the user’s own activity plan, but also included a weather report for the day. Further, it would remind the user of relevant information about television programmes of interest and pass on personalised messages from the experimenter’s team.

At this stage of the system no particular persona or stereotype was established. Because of the restrictive nature of the self-expression capabilities for the user, due to the yes/no buttons, the system’s agent never explicitly self-expressed itself strongly, and the agent never used ‘I’/‘we’, in line with Nass & Brave (2005).

Any persona assigned by the user was an interpretation of what the user perceived the persona to be via its physical appearance, voice, behaviour, dialogue content and how that dialogue was presented. In post-trial interviews and through study of the field experiment output, it could be established that the users

did apply some label to the agent by the end of the iteration. Since this label was not explicitly controlled by the system designer, the label varied from user to user.

An explanation for the labelling of the agent is that the technical limitations are being presented to the user as a type of behaviour and it is this behaviour which is currently not being accounted for. This lack of an account or inconsistencies in the accounting leads to the user having gaps in their knowledge which they fill. These gaps need to be accounted for, and this process is based on stereotypes. If the inconsistencies cannot be accounted for, the jarring of these expectations leads to sanction. This reaction points towards an opportunity to guide the user's perceptions towards a carefully designed persona enforced by consistent behaviour that makes the robot more acceptable and therefore more engaging. A successfully guided stereotype assessment of the user will therefore aid the establishment of a partnership and further guide the expectations of a user, as shown in Eliza and PARRY. To establish the correct persona for an agent requires:

- a guideline for consistency of all external signals of an agent
- a delimiter for the user's expectations
- design that takes into account the need to establish a positive attitude of the user towards the agent

For iteration 2 the aim is to play to the subjects' perceptions from iteration 1. The role of the rabbit continues to manage the activity and fitness of a user through acting as a conduit for information on this topic and the persona must be consistent with this role. To enable consistency throughout the behaviour of the agent, a back story or history of the agent should be established. This story must also fit with the user's perception of the agent, which in this case integrates its physical appearance as a rabbit. This story need not be revealed by the robot through the interactions but allows a means with which to design the dialog and behaviour consistently. One possible design feature is that the rabbit is shy and self-deprecating, which fits with the perception of a small animal and with its role as passing on information when asked. It also contributes to accounting for its behaviour in that it sits in the user's house and does not always talk to the user when they are nearby.

A more proactive persona may intuitively lead to more interactions and more engagement, but that would require more topics for discussion and as noted above, the acceptability of a technology relies on its ability to work accurately and effectively so minimising the scope presents an opportunity to fully develop those topics rather than providing a more widely reaching but limited depth interaction. In addition, this is a novel interface and adding more features adds more for the subject to remember and hence more confusion – directly counter to the often claimed notion that natural language would be a natural and intuitive interface for a machine.

By making our rabbit extremely shy, the subject will be able to account for the rabbit not saying much and, indeed, being hesitant about joining a conversation. There are however limits and a key notion is that the rabbit should not appear rude. To do that, it needs to respond when addressed although the response might be minimal. The aim is to enable the rabbit to recognize when it is addressed and provide a minimal response.

5.1 Back-channel

When engaged in a conversation the person *not* talking participates by nodding his or her head and making appreciative noises. In Japan the expression “hai” is often treated as meaning yes, but in linguistic circles is often translated as “yes, I have heard you, nothing more”. This model of how language works suggests that the purpose of such signaling is to indicate that there is a commitment to engage. That is, in order for a listener to bother *accounting for* some apparently irrelevant action on the part of the speaker, he or she needs to be engaged (as mentioned in Section ??) and this back-channeling signals that engagement. On hearing “hai”, I know that I am being attended to. Such back channelling is semantics free and as such it is feasible that we can re-produce it with our system.

While more and more research goes into social artificial agents, what is missing such studies of back-channel and its significance. From footage of interactions in iteration 1 it could be observed that long pauses between button press and activation of the agent resulted in insecurity and impatience in the user. Further, since a non-responsiveness which leads to impatience may be outside of accountability, it stands to reason that such missing backchannel would also lead to mistrust and alienation of the system to the user. In order to be seen as a social actor, the agent should therefore acknowledge any interaction that is aimed at it. Humans often use smiles, eye blinking, nodding, or small noises such as “Mhm” or “mmm”. The nabaztag cannot change its visual expression, however it does have moving ears, and some flashing lights at its belly, which would be used to define visual feedback. Noises which can be used are freely choosable.

The problem with both streams, visual and oral, are following: since the facial features are not usable, new signals have to be introduced. Such new signals may not be easily understood as back-channeling by the user. Such problem does not exist with the oral feedback and this is the route we intend to take. However, it should be noted that if back-channel messages are just slightly off-timing or wrong, the result in an interruption or competition for the conversational floor (Young & Lee, 2004).

5.2 “Easter eggs”

As mentioned earlier, it would of course be more interesting to have a social agent which is more versed in the real world, but such an approach requires the encoding of vast quantities of common-sens knowledge. This is not only time consuming, projects such as

the CYC project indicate that it is not even possible. This means that efforts have to be focussed on producing islands of knowledge. This suggests that we should concentrate on encoding knowledge relevant to the primary functions of the agent, however the persona of a shy rabbit also enables the idea that our rabbit is fanatical about some very narrow domain for which we can indeed encode a comprehensive amount of knowledge. These specialist domains might not be made explicit and indeed, like in the games industry where the notion of an “easter egg” is common, the rabbit’s expert knowledge may be hidden and left to be discovered. Such a feature would provide variation, and some positive novelty. As an example consider the possibilities if the rabbit was provided with a (percieved) fanatical interest in “Coronation Street”, which is a long-running UK television series. If the user mentions a keyword related to the show such as “Dan” or “kill”, the shy rabbit suddenly starts a conversation about the latest happenings on the soap. Such obsessions would provide a means of providing novel information to the subject without committing the rabbit to an undersanding of the world at large.

6 Conclusions

In this paper we have addressed the problem of robot companion dialogue. We have drawn on the methodology of conversation analysis to illustrate what we believe to be a major deficiency in many current approaches to human-machine dialogue. Conversation analysis can be used to characterise utterances in talk as either seen but unnoticed, noticed and accounted for, or ones that risk sanction. It is argued that accounting for is ones of the means for humans to effortlessly deal with trouble in talk. From the analysis of compute-human dialogues it is evident that trouble can occur too often when the robot produces turns that cannot be accounted for by the human conversation partner, and therefore lead to sanction.

The work of the SERA project, which we have described, aims to address the notion of social engagement and persona in robots and agents. To this end we have begun the iterative development of a robot that can help its user pursue a healthier lifestyle through regular exercise. In this development process we are examining the interactions that users have with the robot over a series of 10-day periods, during which time the robot is in their home. This approach enables us to examine in depth how user’s reactions to, and interaction with the robot change over the period of their relationship.

By using an iterative process, we have been able to examine the expectations of users after the first iteration to ensure that the changes we make can be led by the users. Moreover, this insight means that the system’s dialogue can evolve in a way that should not violate or challenge these expectations.

7 Acknowledgments

The research leading to these results has received funding from the European Community’s Seventh

Framework Programme [FP7/2007-2013] under grant agreement no. 231868.

References

- [Bandura, 1977] Bandura, A. 1977. Self-efficacy: toward a unifying theory of behavioural change. *Psychological review*, **84**, 191–215.
- [de Angeli, 2005] de Angeli, Antonella. 2005 (September). Stupid computer! abuse and social identity. *In: Angeli, Antonella De, Brahnam, Sheryl, & Wallis, Peter (eds), Abuse: the darker side of human-computer interaction (interact ’05)*. <http://www.agentabuse.org/>.
- [Eggins & Slade, 1997] Eggins, Suzanne, & Slade, Diana. 1997. *Analysing casual conversation*. Wellington House, 125 Strand, London: Cassell.
- [Lewin *et al.* , 2005] Lewin, B., Pattenden, J., Ferguson, J., & Roberts, H. 2005. *The heart failure plan*. London: The British Heart Foundation.
- [Prochaska & Velicer, 1997] Prochaska, J., & Velicer, W. 1997. The transtheoretical model of behaviour change. *American journal of health promotion*, **12**, 38–48.
- [Seedhouse, 2004] Seedhouse, Paul. 2004. *The interactional architecture of the language classroom: A conversation analysis perspective*. Blackwell.
- [Skantze, 2007] Skantze, Gabriel. 2007. *Error handling in spoken dialogue systems - managing uncertainty, grounding and miscommunication*. Ph.D. thesis, Department of Speech, Music and Hearing, KTH.
- [Wallis, 2005] Wallis, Peter. 2005 (September). Robust normative systems: What happens when a normative system fails? *In: Angeli, Antonella De, Brahnam, Sheryl, & Wallis, Peter (eds), Abuse: the darker side of human-computer interaction (interact ’05)*. <http://www.agentabuse.org/>.
- [Wallis, 2008] Wallis, Peter. 2008. Revisiting the DARPA communicator data using Conversation Analysis. *Interaction studies*, **9**(3).
- [Wallis *et al.* , 2001] Wallis, Peter, Mitchard, Helen, O’Dea, Damian, & Das, Jyotsna. 2001. Dialogue modelling for a conversational agent. *In: Stumpton, Markus, Corbett, Dan, & Brooks, Mike (eds), Ai2001: Advances in artificial intelligence, 14th australian joint conference on artificial intelligence*. Adelaide, Australia: Springer (LNAI 2256).