


Please cite the Published Version

Wallis, P and Edmonds, B  (2019) How language works & What machines can do about it. In: CUI 2019: 1st International Conference on Conversational User Interfaces, 22 August 2019 - 23 August 2019, Dublin Ireland.

DOI: <https://doi.org/10.1145/3342775.3342804>

Publisher: Association for Computing Machinery

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/628657/>

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

How Language Works & what machines can do about it

Peter Wallis

Centre for Policy Modelling, Manchester Metropolitan University Business School
Manchester, M15 6BH, UK

pwallis@acm.org

CCS CONCEPTS

• **Human-centered computing** → **Interaction design theory, concepts and paradigms**; *Systems and tools for interaction design*.

KEYWORDS

Spoken language interfaces; conversational AI; dialog design; Conversation Analysis; ethnomethods; intentional interface

ACM Reference Format:

Peter Wallis. 2019. How Language Works & what machines can do about it . In *Proceedings of First International Conference on Conversational User Interfaces (CUI'19)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.475/123.4>

M/C: G'day mate.
JUDGE: G'day. How's it going?
M/C: My mother once said "How is it going is a mystery."
She was a bit nuts.
JUDGE: Uh-huh.
M/C: 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?
M/C: Oh well, that's a pity!
JUDGE: I'm sorry; what's a pity?
M/C: Apologies accepted.
JUDGE: I feel compelled to insult you, and we've only been talking a minute.
M/C: We are bloody well talking, aren't we?
JUDGE: Oh baby. We sure are. So here are some keywords for you: shit mother baseball.
...

Figure 1: The 1995 Loebner Prize winner having a bad run.

1 ABSTRACT

The current skills shortage in dialog system development is being filled by very clever graduates that are however victims of historical forces. The current batch of engineers, at all

CUI'19, August, 2019, Dublin, Ireland

© 2019 Association for Computing Machinery.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of First International Conference on Conversational User Interfaces (CUI'19)*, <https://doi.org/10.475/123.4>.

levels, need to study history rather than repeat it. The claim is that techniques from the human and social sciences are the way forward if we are to make better conversational user interfaces.

2 HISTORY

In 1987 James Allen wrote the book on natural language understanding [1] and gave a detailed overview of what we knew then about processing language with a computer. The approach described has been glossed as "The conduit metaphor" [9] and characterised as Meaning-Text Theory [8] by researchers in the former Soviet Union. The gist is that meaning is *in* the text and that by passing utterances through a set of transforms – classically, speech-to-text, morphology, syntax, semantics, pragmatics, and "world knowledge," we can produce (a canonical representation of) the meaning – preferably in something like predicate calculus. Having the meaning, the hard work is done and the rest could be left to others. The legacy of this approach is evident today in everything from the applications deemed suitable for a conversational user interface through the peer review process to the way we design our development environments.

Around 1990 there was a collective shift away from the focus on meaning and toward statistical models of language. For dialog, rather than having the machine understand what was said to it, the focus became what should the machine say next? Using a corpus of human conversations, machine learning could build models of conversation that could answer this question. However the data sparsity issue meant that, rather than words, corpora were annotated with Dialog Acts (DAs) and sequences of these rather than words were used as the training data. Ten years ago POMDP [3, 18, 19] were popular and today deep learning is taking a similar approach. The problem here is not with the machine learning, but with the process of annotating the corpora [7, p289]. Although YES_NO_QUESTION may seem like a reasonable category for a speech-act, how language works is not as obvious as it may seem.

3 LANGUAGE IN ACTION

This historical arc has resulted in successful applications only where there is an obvious link between meaning and action. We have speech interfaces to answer questions - the "meaning" is the set of search terms or the SQL query to run against a relational database, and we have command following speech interfaces - "turn on the kitchen light" maps easily to a required action. But consider the human-machine dialog in Figure 1. To put it bluntly, things just fall apart

In conversation, a conversational partner's utterance will go:

- Seen but unnoticed** if the utterance is the second part of an adjacency pair - a recognisable answer to a question; a follow-up greeting, etc
- Noticed and accounted for** if the speaker can figure out why their conversational partner (CP) said what he or she said. For example follow-up questions, excuses, and so on, or
- Risks sanction.** Sanction can manifest as the speaker simply not wanting to converse with the CP through to out-right abuse.

Figure 2: How language actually works

and it is difficult to see how the meaning of what is said causes this.

Rather than meaning, Conversation Analysis [10] or CA, looks at “the work done” by utterances in context. Consider this conversation in a Doctor’s consultation:

Patient: So this treatment, it won’t affect us having kids will it?

Doctor: [silence 3s]

Patient: it will?

Doctor: I’m afraid so ...

The silence does not have meaning but it certainly has an effect. Critically, as a “member of the community of practice” you know exactly why the patient says what he says without recourse to science. In the Ethnomethodological variant of CA [12] introspection can be used to access this “folk knowledge” about what is done, but it has no part in scientific explanation. However obvious YES_NO_QUESTION may be as a category of speech-act, categories require theory and theory needs grounding in empirical data. Over the last 50 years CA has been refining that process of looking at data and Figure 2 is a summary from Seedhouse [11] of “the findings of CA”. These three stages are demonstrated admirably in Figure 1. Briefly [15], at line 5 the Machine cannot account-for the Judge’s preceding utterance and says so. The Judge explains why he said what he said and, at line 7, moves on. The machine then reacts to the first part (line 6) which does not make any sense to a human because any human understands the explanation and is no longer attending to that issue. The Machine fails to do this. From that point the dialog unravels, the judge gets frustrated, and sanctions his Conversational Partner with a swear word.

Language works, not by passing facts, but by cooperatively developing common ground with participants actively accounting-for their conversational partner’s utterances. What is more, people are compelled to cooperate [13], and expected to work hard at it. Consider this extract from Eggins and Slade talking about sequential relevance:

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 [4].

If computers are going to participate in the human activity we call conversation, they need to be able to “account for” in a way that human interlocutors expect. This requires (something like) intention recognition. The alternative is that voice user interfaces are just a rather poor version of a GUI.

4 THE FUTURE

Existing techniques from the field of NLP handle the “seen but unnoticed,” and the nature of “risks sanction” can be studied using CA, but handling the noticed and accounted for would seem to require full scale automated intention recognition. This is hard [5, 6]. Our epiphany has been to note that there is no point having a system recognise the intent of a user when it does not have a strategy for responding appropriately.

We have been developing a dialog manager based on the principles of behaviour based robotics [2] that tags each behaviour with the goal(s) it might achieve [14]. As an approximation of intention recognition we assume that, if a behaviour is successfully coping with the user’s input, then the user’s intent is the goal associated with the active behaviour. If the behaviour starts to not match input, then we assume the user has a different intent and look for another behaviour (and goal) that does match. Critically, people happily talk about their goals [17] and so the process of changing (shared) goal can be just another behaviour.

Taking an intentional approach, we have had some impressively positive results. Using best practice as the control, we trialed a system which did not give better results – task completion was pegged at 20% – but which participants scored significantly better (indeed positive) on user satisfaction [16]. The explanation for our success is that users “read off” the system’s intent as *wanting* to help. Unsurprisingly, they liked it more. Talk of intent is taboo in some circles however we claim CA(EM) is a methodology that enables us to study intent in a HCI context with some level of scientific rigour and, we claim, is the best way forward if we are to engineer better conversational user interfaces.

5 SUMMARY

This paper has crossed the full spectrum from theory, through methodology, to implementation, none of it thoroughly. There is certainly far more to say, although it has to a large extent all been said before. Progress with CUI could do with a stronger focus on the history rather than vanishingly small incremental improvements in measurable ephemera of a market driven vision of the future.

REFERENCES

- [1] James Allen. 1987. *Natural Language Understanding*. The Benjamin/Cummings Publishing Company, inc.
- [2] Ronald C. Arkin (Ed.). 1998. *Behavior-Based Robotics*. MIT Press, Cambridge, MA.
- [3] Paul A. Crook, Simon Keizer, Zhuoran Wang, Wenshuo Tang, and Oliver Lemon. 2014. Real user evaluation of a POMDP spoken dialogue system using automatic belief compression. *Computer Speech and Language* 28, 4 (2014), 873–887. <https://doi.org/10.1016/j.csl.2013.12.002>
- [4] Suzanne Eggins and Diana Slade. 1997. *Analysing Casual Conversation*. Cassell, Wellington House, 125 Strand, London.
- [5] Clinton Heinze. 2003. *Modelling Intention Recognition for Intelligent Agent Systems*. Ph.D. Dissertation. Department of Computing and Information Systems, University of Melbourne.
- [6] Barbara J. Grosz Karen Lochbaum and Candace Sidner. 1999. Discourse Structure and Intention Recognition. In *A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text*, R. Dale, H. Moisl, and H. Somers (Eds.). Marcel Dekker Inc., Madison Avenue, New York.
- [7] Stephen C. Levinson. 2000. *Pragmatics*. Cambridge University Press, Cambridge, UK.
- [8] I. Mel'cuk. 1981. Meaning-Text Models: a recent trend in soviet linguistics. *Annual Review of Anthropology* 10 (1981), 27–62.
- [9] Michael J. Reddy. 1993. The conduit metaphor: A case of frame conflict in our language about language. In *Metaphor and Thought*, Andrew Ortony (Ed.). Cambridge University Press.
- [10] H. Sacks. 1992. *Lectures on Conversation (edited by G. Jefferson)*. Blackwell, Oxford.
- [11] Paul Seedhouse. 2004. *The Interactional Architecture of the Language Classroom: A Conversation Analysis Perspective*. Blackwell.
- [12] Paul ten Have. 1999. *Doing Conversation Analysis: A Practical Guide (Introducing Qualitative Methods)*. SAGE Publications, California, USA.
- [13] Michael Tomasello. 2008. *Origins of Human Communication*. The MIT Press, Cambridge, Massachusetts.
- [14] Peter Wallis. 2004. Mixed Initiative Dialog using Goal Tagged Activities. In *Embodied Conversational Agents: Balanced Perception and Action (Proceedings of the Workshop at AAMAS'04)*, Catherine Pelachaud, Zsófia Ruttkay, and Kris Thórisson (Eds.). New York, NY.
- [15] Peter Wallis. 2005. Robust Normative Systems: What happens when a normative system fails?. In *Abuse: the darker side of human-computer interaction*, Sheryl Brahnam Antonella de Angeli and Peter Wallis (Eds.). Rome.
- [16] Peter Wallis, Keeley Crockett, and Clare Little. 2014. When Things Go Wrong. In *Human-Agent Interaction Design and Models (HAIDM)*, Sarvapali D. Ramchurn, Joel Fisher, Avi Rosenfeld, Long Tran-Thanh, and Kobi Gal (Eds.). Paris.
- [17] Peter Wallis, Helen Mitchard, Damian O'Dea, and Jyotsna Das. 2001. Dialogue Modelling for a Conversational Agent. In *AI2001: Advances in Artificial Intelligence, 14th Australian Joint Conference on Artificial Intelligence*, Markus Stumptner, Dan Corbett, and Mike Brooks (Eds.). Springer (LNAI 2256), Adelaide, Australia.
- [18] Jason D. Williams and Steve Young. 2007. Partially observable Markov decision processes for spoken dialog systems. *Computer Speech and Language* 21 (April 2007), 393–422. Issue 2.
- [19] S. Young, M. Gasic, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu. 2010. The Hidden Information State Model: a practical framework for POMDP-based spoken dialogue management. *Computer Speech and Language* 24, 2 (2010), 150–174.